1 Introduction to Protein Purification Strategies

Richard J. Simpson,* Tony Pawson,† Gerald Gish,† Peter Adams,‡ and Erica A. Golemis‡

**Joint ProteomicS Laboratory (JPSL) of the Ludwig Institute for Cancer Research and the Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia; †Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada; ‡Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111*

If the structure of a protein cannot yet be accurately predicted from its sequence of amino acid residues, how can we reach an acceptable level of detailed knowledge about its function? For the immediate future, at least, the answer must come from the hard graft of purifying target proteins to a sufficient degree and in enough quantity to allow detailed biochemical and structural analyses. If the investigator is lucky, purification may follow a well-documented protocol. More often than not, however, purification will require modification of an existing protocol or development of an entirely new one. This chapter outlines the factors to be considered in designing a purification protocol and some of the new and exciting trends in protein purification. Much of the theory and practical details of the various separation modalities are considered in later chapters.

THE CHALLENGE

The challenge of protein purification becomes self-evident when one considers the complex mixture of macromolecules present in a biological matrix such as a cell or tissue extract. This complexity is illustrated dramatically by the protein profile of human colonic epithelial cells, determined by two-dimensional gel electrophoresis, as shown in Figure 1.1. In addition to the protein of interest (target protein), several thousand other proteins with different properties are present in any given cell type (a conservative estimate is ~5000–8000), along with nonproteinaceous

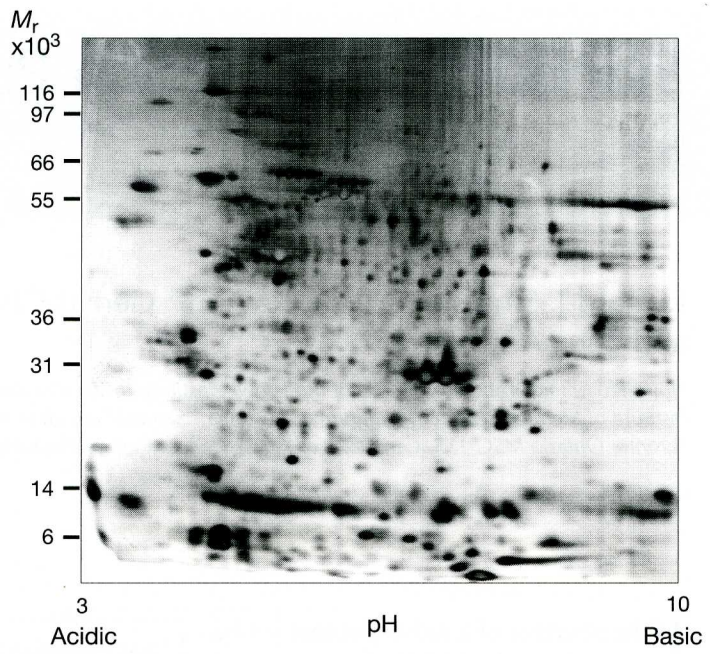


FIGURE 1.1. Human colonic epithelial cell proteins resolved on two-dimensional gels. The approximate pH and apparent molecular weight values are indicated on the *x*- and *y*-axis, respectively.

materials such as DNA, RNA, polysaccharides, and lipids. Proteins are also present within cells in varying amounts. A highly abundant cytoskeletal protein such as actin may constitute 10% of the total weight of protein in the cell extract. At the other extreme, a rare transcription factor may be expressed at such low levels (<0.001%) that only a few molecules are present in each cell. The challenge is to therefore purify (or "capture") the target protein from crude or clarified source material, free of contaminating protein, with reasonable efficiency, speed, yield, and purity.

HOW MUCH PROTEIN IS NEEDED, AND WHAT LEVEL OF PURITY IS REQUIRED?

Just how extensively a protein must be purified, as well as the scale of the purification, depends largely on the end goals, i.e., the purpose for purifying a protein in the first place. In general, one would like to obtain a target protein in a homogeneous form, free of all contaminating proteins and other materials. However, there are many applications where less than completely pure protein will suffice. The scale of protein purification depends on the amount of material required to perform a particular task. A brief outline of the quantities of protein and degree of purity required for many protein applications is given in Table 1.1.

Normal laboratory-scale biochemical purification, using conventional chromatography equipment, will typically produce 1–100 mg of protein. Microscale purification, which produces nanograms to micrograms of protein, requires chromatographic equipment specifically designed for the purpose, in particular, liquid chromatographs capable of delivering low flow rates and accurate elution gradients (Simpson and Nice 1989; Nice and Catimel 2000).

At the extreme end of the scale, if the goal is to obtain limited amino acid sequence information for the purposes of identifying an unknown protein, then only a few micrograms (1–5 pmoles) of highly purified material are required for "state-of-the-art" protein/peptide sequencing methods employing automated Edman degradation instruments. For an excellent review of the chemistry involved in Edman degradation (Edman 1949), an appraisal of the various instruments,

TABLE 1.1. Quantity and purity of protein required for different applications

Application	Amount Required	Purity Required	Comments
Identification	0.002–0.2 µg	high (>95%)	Amino-terminal sequence analysis of proteins using the Edman degradation procedure requires 5–10 pmoles (see Chapter 6 in Simpson 2003), whereas mass spectrometric approaches that rely on protein identification by accurate measurement of peptide masses and/or sequencing by collision-induced dissociation of peptides (e.g., tandem mass spectrometry) require 0.2–1 pmole of peptide(s) (see Chapters 7 and 8 in Simpson 2003). Both methods are destructive and the sample cannot be recovered.
Immunology *polyclonal* antibodies monoclonal antibodies protein microarrays	µg–mg	medium-high	Tens of micrograms of protein may be required as an immunogen during the preparation of polyclonal antisera and monoclonal antibodies (see Harlow and Lane 1999). The higher the purity of the immunogen, the greater the chance of raising an antibody response of high specificity. Depending on the state of the immunogen and the mode of immunization, the antibody may react only with the native form of the target protein, only with denatured forms, or with both. Antibodies that react with denatured forms are capable of detecting nanogram amounts of target protein in, for example, western blots and protein microarrays.
Enzymology	1–5 mg	high (>95%)	The amount of target protein required for enzymological studies depends on the sensitivity of the assay. The degree of purity required depends on the specificity of the assay and its susceptibility to interference by contaminants. Because these assays generally need to be repeated many times and are destructive, a reasonable goal might be to purify 1–5 mg of the target protein to >90% purity.
Biophysical analysis	mg–g	high (>95%)	Many of the methods used for biophysical characterization of target proteins (e.g., fluorescence, UV absorption spectroscopy, analytical ultracentrifugation, surface plasmon resonance, and CD analysis) permit recovery of the sample for further use.
Three-dimensional structure	10–20 mg	high (>95%)	X-ray crystallography requires ~1–2 mg of target protein to establish conditions for crystallization. An additional 5–10 mg is needed to grow crystals, large enough for X-ray diffraction. Preferably, protein should be of the highest purity achievable. Initial screening attempts can be made using target protein of ~80% purity. If no crystals are obtained, then higher purity is essential. NMR requires ~0.5 µmole of target protein (10 mg for a 20-kD protein) for initial spectra. Typically, $^{15}N/^{13}C$-labeled protein is required to solve the structure of proteins (5–20 kD). Although prolonged exposure of crystals to X-rays during collection of diffraction data may cause radiation damage to the protein, both NMR and X-ray crystallography are essentially nondestructive methods and material can be recovered for other tasks.
Pharmaceutical	mg–kg	high (>99.9%)	For clinical use, pharmaceutical proteins must be free of pyrogens and bacterial endotoxins and stable upon extended storage.

and requirements for preparation of samples, see Lottspeich et al. (1994). A detailed discussion of current automated chemical amino acid sequencing procedures (both amino-terminal and carboxy-terminal methods) including a sample preparation are presented elsewhere (Chapter 6 in Simpson 2003). With careful optimization of current instrumentation, sensitivities on the order of 0.1–1 pmole are achievable using sequencing cycles of 20 minutes per residue (Henzel et al. 1999).

One of the most exciting innovations in protein chemistry during the past decade is the development in mass spectrometry (MS) of new "soft ionization" techniques, such as electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI). These modern MS-based techniques require only a few nanograms (0.2–1 pmole) of material to identify proteins (and peptides) on the basis of peptide masses and peptide fragment ion data, which are generated by collision-associated dissociation (CAD) of peptide ions (tandem MS). One corollary of MS-based protein identification is that the criterion of protein purity required has changed profoundly,

being much less stringent than that required for conventional Edman degradation. Indeed, with MS methods, it is now possible to identify proteins in complex mixtures without prior fractionation of the mixture (see Chapter 8 in Simpson 2003).

If the goal of the purification is to obtain enough enzyme for detailed physical and kinetic studies, milligram quantities of highly purified protein are necessary. Comparable quantities of target protein are required for performing biophysical studies of protein stability and kinetics of folding that involve techniques such as fluorescence (and ultraviolet [UV] absorption) spectroscopy, circular dichroism (CD), differential scanning calorimetry, analytical ultracentrifugation, surface plasmon resonance, and hydrogen exchange.

If the purpose of purifying a protein is to determine its three-dimensional structure, either by X-ray crystallography or by nuclear magnetic resonance (NMR), tens of milligrams of highly purified material may be required. The uncertainty in the amounts required to carry out a full chemical and physical analysis of a protein using these techniques is due to the variable behavior of proteins in solution, particularly at the high concentrations (8–10 mg/ml) required for analysis. Although some proteins remain in their monomeric form over a broad concentration range (which is desirable for NMR analysis), others will aggregate or even precipitate. For X-ray diffraction analysis, a further dilemma lies in what has become a primary problem for this technology—that of growing large, single crystals (McPherson 1997). Glycosylated proteins, in particular, present a major problem due to their carbohydrate-associated charge heterogeneity and the potential of the oligosaccharides to inhibit protein–protein contacts necessary for crystal lattice packing. Biochemists are now attuned to this problem and have developed the expertise to reduce this heterogeneity by treating the purified protein with exoglycosidases or, alternatively, by using mammalian expression systems (e.g., Chinese hamster ovary cells) with genetically altered carbohydrate biosynthesis pathways for the production of recombinant glycoproteins with minimal carbohydrate heterogeneity (Stanley 1989). Investigators must be careful here and not deplete the carbohydrate completely, because fully deglycosylated proteins are often difficult to solubilize, especially at the high protein concentrations required for biophysical studies. Typical requirements for the crystallization of proteins are 5–10 mg of target protein (8–10 mg protein/ml) for screening studies and for crystallization. For reviews on protein crystallization methodologies, see McPherson (1990, 1997) and Shu and Bi (1997). In the case of NMR analysis of peptides and proteins, sample purity and the choice of solution conditions are critical factors. Samples need to be chemically homogeneous and free from low-molecular-weight protonated molecules, such as those used in many of the common biological buffers (e.g.,

CONVERTING MOLES TO MICROGRAMS

Protein molecular mass (daltons)	Amount of protein per nmole	Number of moles in 1 μg of protein
1,000	1 μg	1 nmole or 6×10^{14} molecules
10,000	10 μg	100 pmoles or 6×10^{13} molecules
20,000	20 μg	50 pmoles or 3×10^{13} molecules
50,000	50 μg	20 pmoles or 1.2×10^{13} molecules
100,000	100 μg	10 pmoles or 6×10^{12} molecules
200,000	200 μg	5 pmoles or 3×10^{12} molecules

One mole of a protein is the amount that contains 6.023×10^{23} molecules of that protein, which is known as Avogadro's number. The weight of a mole of a protein in grams (g) is the same as its molecular mass. For example, for a protein with a molecular mass of 20,000 daltons, the weight of 1 mole of the protein is 20,000 g.

1 mmole = 10^{-3} moles
1 μmole = 10^{-6} moles
1 nmole = 10^{-9} moles
1 pmole = 10^{-12} moles
1 fmole = 10^{-15} moles

Tris, MOPS, and TEA). A typical solvent used in NMR studies is H_2O (and dimethylsulfoxide, methanol, and acetonitrile for peptides only). For a review on the practical considerations of NMR spectroscopy of peptides and proteins, see Hinds and Norton (1997).

USING RECOMBINANT PROTEINS OFTEN SIMPLIFIES THE PURIFICATION PROCESS

Another important consideration in designing a purification strategy is whether the investigator is attempting to purify a native protein from a biological matrix (e.g., a cell line, cell-conditioned medium, or tissue) and relying on a biological assay or antibody for monitoring the purification, or whether the task is to purify an overexpressed recombinant protein. Purifying rare proteins (e.g., growth factors, receptors, or transcription factors) from natural biological sources is often extremely difficult, requiring extremely large quantities of starting material and a 1–2-million-fold purification to achieve homogeneity (see Table 1.2).

In contrast, purifying overexpressed recombinant proteins in milligram to kilogram quantities (especially for pharmaceuticals) has been greatly simplified by the ability to produce target proteins containing a fusion partner (or "a purification handle") designed to facilitate protein purification. Biotechnology companies have commercialized a remarkable variety of sophisticated fusion proteins available for biological research. The utility of fusion proteins in an increasing range of applications is examined in a series of comprehensive reviews by Makrides (1996) and Sambrook and Russell (2001).

TABLE 1.2. Examples of low-abundance proteins and peptides isolated from natural biological sources

Protein	Source	Yield (μg)	Reference
Multipotential colony-stimulating factor	pokeweed mitogen-stimulated mouse spleen-cell-conditioned medium (10 liters)	1	Cutler et al. (1985)
Human A33 antigen	human colon cancer cell lines (10^{10} cells)	2.5	Catimel et al. (1996)
Platelet-derived growth factor (PDGF)	human serum (200 liters)	180	Heldin et al. (1981)
Granulocyte-colony-stimulating growth factor (G-CSF)	mouse lung-conditioned medium (3 liters)	40	Nicola et al. (1983)
Granulocyte-macrophage colony-stimulating growth factor (GM-CSF)	mouse lung-conditioned medium (3 liters)	12	Burgess et al. (1986)
Coelenterate morphogen	sea anemone (200 kg)	20	Schaller and Bodenmuller (1981)
Peptide YY (PYY)	porcine intestine (4000 kg)	600	Tatemoto (1982)
Tumor necrosis factor (TNF)	HL60 tissue culture medium (18 liters)	20	Wang and Creasy (1985)
Murine transferrin receptor	NS-1 myeloma cells (10^{10} cells)	20	van Driel et al. (1984)
Fibroblast growth factor (FGF)	bovine brain (4 kg)	33	Gospodarowicz et al. (1984)
Transforming growth factor-β (TGF-β)	human placenta (8.8 kg)	47	Frolik et al. (1983)
Human interferon	human leukocyte-conditioned medium (10 liters)	21	Rubinstein et al. (1979)
Muscarinic acetylcholine receptor	porcine cerebrum (600 g)	6	Haga and Haga (1985)
β_2-adrenergic receptor	rat liver (400 g)	2	Graziano et al. (1985)

Adapted, with permission, from Simpson and Nice (1989).

The growing use of recombinant proteins as pharmaceuticals requires extraordinarily high levels of purity (>99.99% purity in some cases) in order to remove materials that would be harmful when injected into a human patient (e.g., pyrogens and viruses).

PROTEINS CAN BE SEPARATED ON THE BASIS OF THEIR INTRINSIC PROPERTIES

Many proteins and peptides of biological interest are of very low abundance, often constituting <0.1% of the total cellular proteins (e.g., growth factors and signal-transducing molecules). Hence, the following are obvious constraints in obtaining sufficient quantities of such proteins in a homogeneous form suitable for biological testing and identification purposes:

- The large quantities of source material required (for examples, see Table 1.2).
- The availability of separation facilities (e.g., instrumentation).
- The physical constraints of chromatographic resin support capabilities (i.e., capacity and flow rates).

To fully exploit the chemical and physical properties of a target protein in designing an appropriate strategy for its purification, the following parameters for the target protein should be obtained in initial pilot studies:

- molecular weight (e.g., by SDS-PAGE, size-exclusion chromatography, or analytical ultracentrifugation)
- pI (e.g., by isoelectric focusing)
- stability with respect to pH, salt, temperature, proteases, inclusion of additives to protein solvents to maintain biological activities (e.g., detergents, thiol reagents, and metal ions).

Given that the logistical problems can be solved, the tremendous variation in physical and chemical properties among proteins can usually be exploited to design a workable purification scheme. Some of the relevant properties of proteins and how they can be used to purify proteins are described in the remainder of this section.

Exposed Amino Acid Side Chains Determine Protein Solubility

The solubility of a protein can differ markedly depending on the solvent, and different proteins in the same solvent can vary greatly from one another in their solubility. This protein-to-protein variation in solubility is due to the differences in the ratio of solvent-exposed charged (i.e., polar) and hydrophobic amino acids on protein surfaces. Parameters that influence the solubility of a protein include solvent pH, the ionic strength and nature of the buffer ions, solvent polarity, and temperature. Because it is not possible to predict with accuracy the solubility properties of proteins, much of the skill in purification comes from experience in handling proteins under a variety of conditions.

Proteins tend to precipitate differentially from aqueous solution upon the addition of neutral salts (e.g., ammonium sulfate), polymers such as polyethylene glycol, or organic solvents (e.g., ethanol or acetone). This behavior provides a simple means to concentrate proteins at high yield from large volumes with a concomitant two- to three-fold degree of purification. It is therefore very common for this technique to be used at an early stage of a purification, typically at the stage immediately following solubilization of the starting tissue, particularly when working on a moderate to large scale, where sample volumes are often large.

In some cases, a target protein is known to display an *unusual property* that can be exploited in a purification strategy. For example, incubation of a crude extract at low (or high) pH might lead to selective precipitation of the majority of proteins, which can be readily removed by centrifuga-

tion (because most proteins denature and precipitate at extremes of pH), leaving the target protein in the supernatant in a more highly purified form. The high stability of muscle adenylate kinase at pH 2 has been used to advantage during the purification of this enzyme (Heil et al. 1974).

The Size and Shape of Proteins Affect Their Movement Through Liquids and Gels

The size and shape of a protein affect its ability to move in a fluid solution. For a review of this subject and the variety of techniques available to estimate the molecular weight and shape of macromolecules, see Cantor and Schimmel (1980). Proteins vary markedly in size, ranging from a few amino acid residues (e.g., peptides) of a few hundred daltons to more than 3000 amino acids with a molecular mass in excess of 300,000 daltons. However, the molecular mass of most proteins falls in the range of 6 kD to 200 kD (see Fig. 1.1). Whatever its mass, the shape of a protein, which influences its movement through fluids, ranges from nearly spherical (globular) to markedly asymmetric.

These properties of proteins are exploited directly in the purification techniques of size-exclusion chromatography (SEC; Chapter 6). In SEC, a protein solution is passed through a column of porous beads (a large range of beads with variable, but defined, pore sizes are commercially available). The internal diameter of the pores are such that large proteins do not have access to the internal space of the bead, whereas small proteins have free access, and proteins of intermediate size have partial access. Large proteins therefore pass directly through the column, and smaller proteins are retarded. Although the capacity of this method is low, and its resolving capability is limited, SEC is useful for separating proteins with extremes in size. Because SEC supports are non-interactive, they have no trace-enrichment potential and cannot be used for the purpose of concentrating proteins.

One of the most important separation techniques in protein chemistry that exploits the size of proteins and peptides is SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (see Chapter 3). Analytical SDS-PAGE has long been used to monitor the purity and recovery of a target protein after each step in a purification protocol. However, equipment is now available that enables the method to be used in a preparative manner. In this method, proteins are denatured and fully coated with the negatively charged detergent SDS, such that they migrate in electrophoretic gels on the basis of their molecular weight. The pore size of the gels can be varied by altering the amount of cross-linking agent used. The gel has a sieving effect, so that the smallest proteins migrate most rapidly and the larger proteins migrate more slowly. The extremely high resolving power of this method forms the basis of the second dimension of separation in two-dimensional gel electrophoresis. In the first dimension of separation, electrophoresis is performed in the absence of SDS in a gel in which a pH gradient has been established. At a pH characteristic of each protein (the isoelectric point), the net charge on the molecule is zero and it ceases to migrate. During the past 15 years, methods have been developed for isolating and identifying proteins separated by two-dimensional gel electrophoresis (see Chapter 4). Currently, two-dimensional gel electrophoresis is one of the preferred methods in the burgeoning field of proteomics (i.e., the systematic analysis of cell or tissue protein profiles in normal and diseased states). Although high-resolving SDS-PAGE and two-dimensional gel electrophoresis methods are limited in capacity, and the protein is obtained in a denatured state, they provide a ready means for isolating small amounts of a target protein from a complex mixture.

Differences in the Surface Charge of Proteins Are Exploited in Ion-exchange Chromatography

The net charge on a protein is the sum of the positively and negatively charged amino acid residues, at the pH of the solvent. Proteins with a preponderance of basic amino acids (e.g., arginine, lysine, and histidine), referred to as *basic proteins*, will have a net positive charge at neutral pH. Conversely, proteins rich in acidic amino acids (e.g., aspartic acid and glutamic acid) will have an overall negative charge at neutral pH and are referred to as *acidic proteins*. The pH at which the

net charge of a protein is zero is referred to as the isoelectric point (pI). Differences in surface charge can be exploited to separate proteins using ion-exchange chromatography (see Chapter 5). This method relies on a protein carrying a net charge of one sign binding to a solid chromatographic support bearing charged groups of the opposite sign. Proteins can be eluted from the chromatographic support by exchanging the proteins for buffer ions of the opposite charge. This is accomplished by developing the column with a gradient of increasing ionic strength. Each protein then elutes at a concentration of ionic species determined by the magnitude of the protein's surface charge. This technique, which is of intermediate resolution and high capacity, is highly selective and, in some cases, can resolve two proteins differing in only one charge.

Ligand-binding Proteins May Be Purified by Affinity Chromatography

Most proteins exert their biological function by specifically interacting with some other cellular component. For example, enzymes bind effector molecules such as cofactors, substrates, activators, inhibitors, and metal ions (e.g., Cu^{2+}, Zn^{2+}, Mg^{2+}, and Co^{2+}); hormones bind to receptors; transcription factors bind to nuclear and cytoplasmic locations, export signals, and DNA templates. These specific binding phenomena can be exploited by binding the target protein to a chromatographic column carrying the appropriate ligand or metal ion. Elution is achieved by varying the solvent conditions or introducing a solute that competes for the binding of target protein to the ligand. Various types of chromatography based on protein surface recognition or affinity, generically termed affinity chromatography, are described in Chapter 8.

The unique topology of a protein can be exploited for purposes of purification if an antibody is available that recognizes a specific determinant or epitope on the surface of that protein. Such a determinant may comprise a linear array of amino acids (linear determinant) or, alternatively, amino acids that are distributed widely in the polypeptide chain, but come into close spatial proximity upon folding of the native protein (conformational determinant). The exquisite specificity of antibody–antigen interactions forms the basis of immunoaffinity chromatography where a monospecific antibody affixed to a column selectively captures the protein of interest (see Chapter 8).

In recent years, a number of nonbiospecific affinity adsorbents have been developed. These are materials that do not interact with a target protein through a physiological ligand-binding site, but rather with other parts of the protein surface (Scopes 1987; Lowe et al. 1992). Principal among these are (1) the triazine dye ligands (Turner 1981; Qadri 1985) that have been found, by trial and error, to be highly specific for certain proteins and (2) metal atoms attached to a chromatographic support (immobilized metal affinity chromatography, IMAC) (Sulkowski 1985; see Chapter 9). IMAC takes advantage of amino acids such as histidine that can act as electron donors and thereby chelate a metal ion, thus preferentially detaining the target protein on the IMAC column. IMAC using chelated nickel has gained popular acceptance as a method to purify recombinant proteins engineered with a hexahistidine tag at either their amino or carboxyl terminus.

In general, affinity chromatography methods have high resolving capability, but low capacity, and are thus usually reserved for a late stage in a purification protocol. The exception is affinity-based purification of recombinant proteins containing an immunological or other type of affinity "tag" specifically designed to facilitate their purification.

Posttranslational Modifications Provide Additional Opportunities for Purification by Affinity Chromatography

Posttranslational modifications are fundamental to processes controlling cellular behavior, including cell signaling, growth, and transformation (Han and Martinage 1992; Parekh and Rohlff 1997). Many proteins are modified posttranslationally by the addition of carbohydrates to form glycoproteins, phosphates to form phosphoproteins, and lipids to form lipoproteins. For reviews of posttranslational modifications of proteins, see Krishna and Wold (1993, 1997). In many cases, these posttranslational modifications provide recognition "handles" that can be used in protein

fractionation. For example, glycoproteins can be captured from mixtures of proteins by binding them to columns containing immobilized lectins, which are a class of plant proteins capable of selectively binding to particular carbohydrates (see Chapter 8) (Yan and Grinnell 1989; Han and Martinage 1993).

Similarly, many types of phosphoproteins can be captured from complex protein mixtures by binding to a column containing immobilized antibodies directed against phosphotyrosine or, alternatively, using IMAC (see Chapter 9).

Thermostable Proteins Can Often Be Purified Easily

Proteins are typically inactivated and precipitate if heated to 95°C. However, some proteins exhibit a remarkable degree of thermoresistance. Under these circumstances, substantial purification of a target protein can be accomplished by heating a crude extract at a temperature where the target protein is stable (active and soluble), but extraneous proteins are denatured and precipitate from solution. Examples of a purification strategy involving thermostability are those of stathmin, a key mammalian intracellular regulatory protein (Koppel et al. 1990), muscle phosphatase inhibitor-1 (Nimmo and Cohen 1978), and that of *Escherichia coli* alkaline phosphatase. *E. coli* alkaline phosphatase has an additional unusual property, an innate resistance to digestion with proteases. For example, after heat treatment of an *E. coli* cellular extract and removal of the precipitated proteins by centrifugation, the supernatant containing active alkaline phosphatase can be treated with a protease to digest the remaining soluble contaminating proteins, thereby yielding highly purified and active alkaline phosphatase.

DEVISING STRATEGIES FOR PROTEIN PURIFICATION

Before attempting to design a purification scheme, it is always worthwhile to carry out pilot experiments on the crude extract to determine whether the target protein possesses any unusual chemical and physical properties that might be exploited in a purification strategy. Useful information includes approximate molecular weight and pI; degree of hydrophobicity; presence of carbohydrate (glycoprotein); phosphate modification; free sulfhydryl groups; stability with respect to pH, salt, temperature, proteolytic degradation, and mechanical shear; and bioaffinity for heavy metals. If the nucleotide sequence is known, much of this information might be obtained by close inspection of the deduced amino acid sequence; otherwise, it can be obtained from pilot experiments using crude extracts.

Is Retention of Biological Activity Essential?

An important consideration is whether it is essential to retain biological activity of the target protein during purification. Most proteins retain activity at low temperatures in neutral aqueous buffers containing stabilizing additives such as glycerol and detergents. These conditions are incompatible with techniques in which the chromatography conditions are somewhat harsh. For example, RP-HPLC (see Chapter 7) requires the use of organic solvents (e.g., acetonitrile) and ion-pairing acids (e.g., trifluoroacetic acid) to elute proteins from the reversed-phase column. Very few proteins are able to retain their biological activity in the face of such abuse, although there are notable exceptions, such as growth factors. Similarly, the conditions required to elute proteins from immunoaffinity columns are often severe, due to the high binding affinity of antibody/antigen complexes, which dissociate only at extremes of pH (e.g., 10 mM HCl, 10 mM NaOH) or in high concentrations of salt (e.g., 3 M $MgCl_2$) (see Chapter 8).

Unless the target protein has special surface recognition characteristics that can be exploited, for example, biospecific affinity for ligands or nonspecific affinity for triazine dyes, ion-exchange

chromatography, SEC, and HIC are the methods that offer the best chance of retaining biological activity. To limit the losses of biological activity of labile target proteins during purification, it is important to minimize the number of steps in the purification protocol and, where possible, avoid the need for buffer exchange between steps. It is also important to discriminate between losses of biological activity due to denaturation and physical losses caused by irreversible adsorption to the chromatographic support or by proteolytic degradation. Finally, a robust biological assay that is sensitive, and can handle a large number of samples with a rapid turnaround, is an essential requirement of a successful protein purification protocol.

Where retention of biological activity of the target protein is not required (but where the progress of purification can be monitored by western blot analysis or biosensor analysis; Nice and Catimel 1999), there is no restriction on the purification techniques that can be brought to bear on the problem. Foremost among these, especially when the end goal is identification by amino acid sequence analysis, is the high-resolving method of SDS-PAGE (see Chapters 3 and 4), followed by characterization and identification by Edman degradation or methods based on mass spectrometry (see Chapters 6 and 8 in Simpson 2003).

How Many Purification Steps Are Necessary?

Only rarely can a protein be purified to homogeneity in a single step, even when this step is based on an exquisitely specific biological characteristic. According to an analysis of 100 papers describing a successful protein purification, the average number of steps necessary to purify proteins to homogeneity is four, with an overall yield of 28% and a purification factor of 6380, corresponding to an average ninefold purification and 73% yield per step (Bonnerjea et al. 1986). In addition to the purification steps, there is often a need to concentrate and/or clarify the initial cell lysate or tissue extract and, sometimes, to include procedures for exchange of buffers between purification steps. The four generic steps of a typical protein purification protocol are illustrated in Figure 1.2.

It is generally recognized that with most conventional chromatographic supports (packings), there are compromises among speed, resolution, recovery, and capacity. Typically, this relationship is depicted as a trigonal pyramid with each apex labeled with one of these parameters (Fig. 1.3). In practice, there is interdependence between these parameters—for example, an increase in speed of a chromatographic step is usually at the expense of resolution or capacity, or an increase in resolution or capacity is usually achieved at the expense of speed. A list of procedures for fractionat-

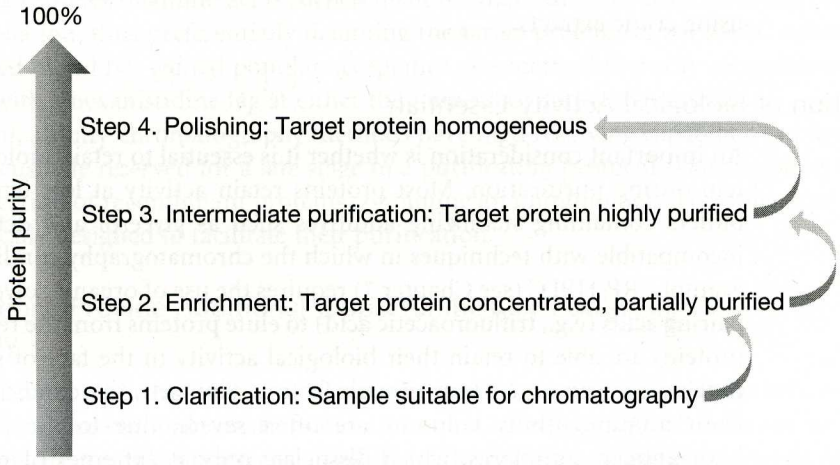


FIGURE 1.2. Successive stages in a typical purification protocol. Most protein purification schemes consist of at least four stages: a preliminary clarification stage, an initial enrichment (or "capture") stage, the intermediate purification, and the final polishing stage to yield a homogeneous target protein.

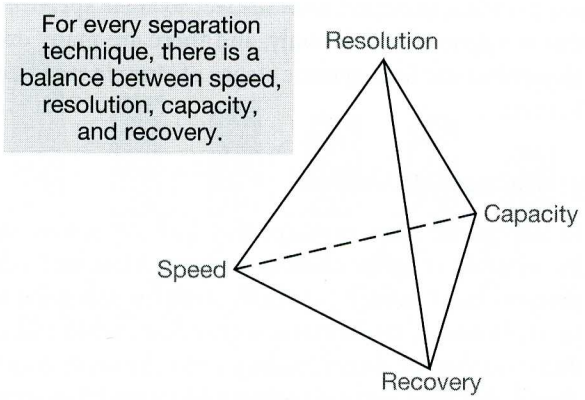
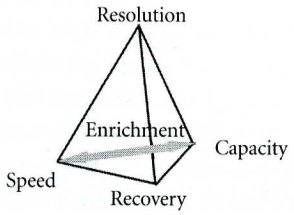
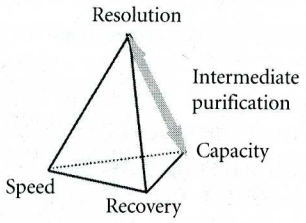


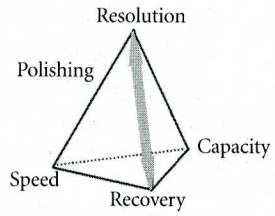
FIGURE 1.3. Model depicting the interrelationships between important chromatographic parameters.

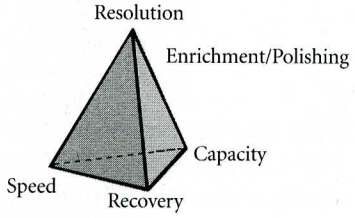
TABLE 1.3. Procedures for fractionating proteins and peptides

Purification Stage	Basis of Separation
NATIVE PROTEINS	
Enrichment (capture) stage: Rapid/high-capacity, low-resolution modes	
Chromatography	
ion-exchange	charge
hydrophobic interaction	hydrophobicity
affinity - DNA	DNA binding
- dye	specific dye-binding affinity
- substrate	ligand binding site
- lectin	carbohydrate content and type
- immobilized metal affinity (IMAC)	metal binding
- immunoaffinity	specific antigenic site
Intermediate purification stage: High-capacity, low-resolution modes	
Chromatography	
ion-exchange	charge
hydrophobic interaction	hydrophobicity
size exclusion	size, shape
chromatofocusing	pI
Electrophoresis	
gel-electrophoresis (preparative gels)	charge, size, shape
Final polishing stage: Low-capacity, high-resolution modes	
Chromatography	
reversed-phase HPLC	hydrophobicity, size
size exclusion	size, shape
Electrophoresis	
gel electrophoresis (analytical scale SDS-PAGE, 2-DE)	charge, size, shape
isoelectric focusing	pI
free-flow electrophoresis	charge, size, shape
RECOMBINANT PROTEINS WITH A FUSION TAIL (OR "TAG")	
Combined enrichment/polishing purification stage: High-capacity, rapid/high-resolution modes	
Chromatography	
affinity - substrate	enzyme/ligand-binding site
- lectin	carbohydrate-binding domain
- immobilized metal affinity (IMAC) (e.g., poly[His])	metal binding
- immunoaffinity	antigenic epitopes
hydrophobic interaction	
- hydrophobic amino acid tails (e.g., poly[Phe])	hydrophobicity
ion-exchange	
- charged amino acid tails (e.g., poly[Arg])	charge (or precipitation)









ing proteins, grouped with respect to their speed, resolution, and capacity, is given in Table 1.3; this is a generalized list only, and there are many exceptions to the rule. The strategy and rationale behind the four generic stages of protein purification, as outlined in Figure 1.2, are dealt with in turn.

Step 1: Clarifying the Starting Material

In any purification protocol that includes chromatographic methods, it is always important to incorporate an initial clarification step. Most methods for obtaining a crude extract of intracellular proteins (see Chapter 2) include steps for removing insoluble residues (e.g., differential centrifugation). However, crude extracts are often turbid and contain lipid droplets, and hence, they are seldom suitable for direct loading onto chromatographic columns, due to the likelihood of causing irreparable damage (e.g., column blockage). Since differential centrifugation (including coarse filtration through a plug of glass wool or fine mesh cloth between centrifugation steps) is awkward when handling large sample volumes and seldom results in a clear extract, it is beneficial to couple the clarification step to a rapid concentration step such as fractional precipitation using salts, polymers, or organic solvents. The most common precipitating agents are shown in Table 1.4. In addition to concentrating the protein bulk in the crude extract (this is particularly important when dealing with large sample volumes), fractional precipitation can also afford a two- to eight-fold purification of the target protein (Fig. 1.4). Precipitation is less prone to interference by nonproteinaceous material than are adsorption or chromatography procedures, and the method has a very high (~80%) average yield. It is always desirable that the initial clarification/concentration step be as rapid as possible to guard against proteolytic degradation of the target protein.

In addition to fractional precipitation, ultrafiltration is widely used for the rapid and gentle concentration of large volumes of starting material (e.g., cell-conditioned medium). Since ultrafiltration membranes are available with a variety of molecular-mass cutoff limits (1000–300,000 daltons), a modest purification of the target protein is often achieved during concentration of proteins from large volumes.

Step 2: Capturing the Target Protein

Following initial clarification/concentration, it is prudent to include an enrichment step provided the sample is amenable to column chromatography. The goal is to enrich the target protein as quickly as possible to avoid the risk of losses due to proteolytic degradation and/or other modifications. For native proteins, enrichment is best accomplished using *high capacity/low resolution* chromatographic procedures (see Table 1.3) such as anion-exchange chromatography (see Chapter 5), HIC, or nonbiospecific affinity chromatography (e.g., triazine dye chromatography). Resist all temptation to use an extraordinarily high-resolving method such as immunoaffinity chromatography as the first chromatographic step. In many cases, the high cost of immunosor-

TABLE 1.4. Procedures for clarifying and concentrating large-volume protein samples

Concentration Process	Basis of Method
Precipitation	
ammonium sulfate (including "salting-out" chromatography)	differential solubility
polyethylenimine	forms insoluble complexes with acidic macromolecules (i.e., acidic proteins, DNA, RNA)
ethanol	differential solubility
Phase-partitioning (e.g., polyethylene glycol)	differential solubility
Ultrafiltration	size and shape

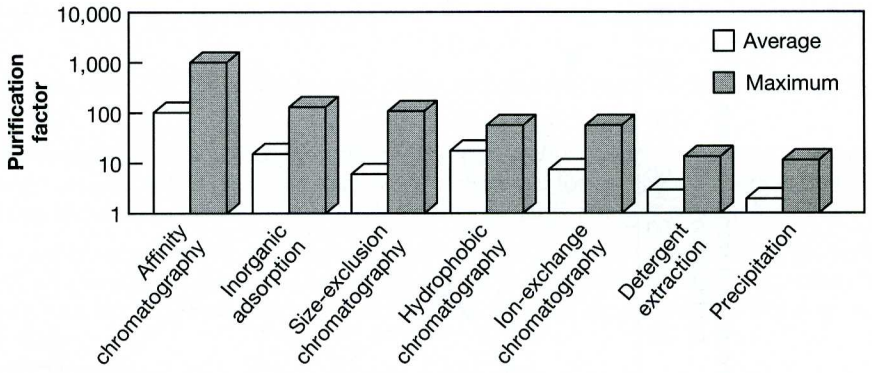


FIGURE 1.4. Average purification factors for various purification methods. (Adapted, with permission, from Bonnerjea et al. 1986.)

bents employing immobilized monoclonal antibodies prohibits the use of large columns. Repeated injections of sample therefore become necessary, which as well as being time-consuming, increase the risk of proteolytic damage to the target protein (and reduce the life of the immunosorbent). Similarly, to choose reversed-phase chromatography as the first step would be inappropriate, due to the low capacity of reversed-phase media and their lack of compatibility with subsequent purification steps, thus requiring a buffer-exchange step.

In the case of recombinant proteins, several fusion systems have been developed to promote their efficient recovery and purification from crude cell extracts or culture media (for reviews, see Ford et al. 1991; LaVallie and McCoy 1995; Makrides 1996). In these systems, a target protein is genetically engineered to contain a carboxy- or amino-terminal fusion tail (or purification "tag") that provides the biochemical basis for some form of affinity chromatography. A variety of fusion tails have been used, including entire enzymes with affinity for immobilized substrates or inhibitors; antigenic epitopes with affinity to immobilized monoclonal antibodies; oligohistidine for recovery by IMAC; carbohydrate-binding proteins or domains recognized by lectins; and a biotin-binding domain for in vivo biotinylation, which creates affinity of the fusion protein to avidin or streptavidin. With some experimentation, fusion tails can be found that do not interfere with the biological activity of target proteins; indeed, in some cases, they have been shown to stabilize it (Ford et al. 1991). If required, a specific protease cleavage site can be engineered into the fusion tail to facilitate removal of the tail using recombinant proteases (Ford et al. 1991; Walker et al. 1994).

Step 3: Purifying and Concentrating–Intermediate Steps

This step should be designed to provide further purification (removal of extraneous protein) and reduction of sample volume (concentration), and it is best accomplished using intermediate capacity/intermediate to high-resolution chromatography (see Table 1.3). For example, SEC can be optimized to yield high resolution (see Chapter 6), but only at low speed and with small sample volumes. Likewise, immunoaffinity chromatography (see Chapter 8) becomes a viable option at this stage of the process.

Step 4: Final Polishing

The purpose of the final polishing step(s) is to remove any minor contaminants remaining (including posttranslationally modified target protein), to remove possible aggregates, and to prepare the homogeneous target protein for its immediate use or for storage. This step is best accomplished using intermediate to low-capacity/intermediate to high-resolution procedures (Table

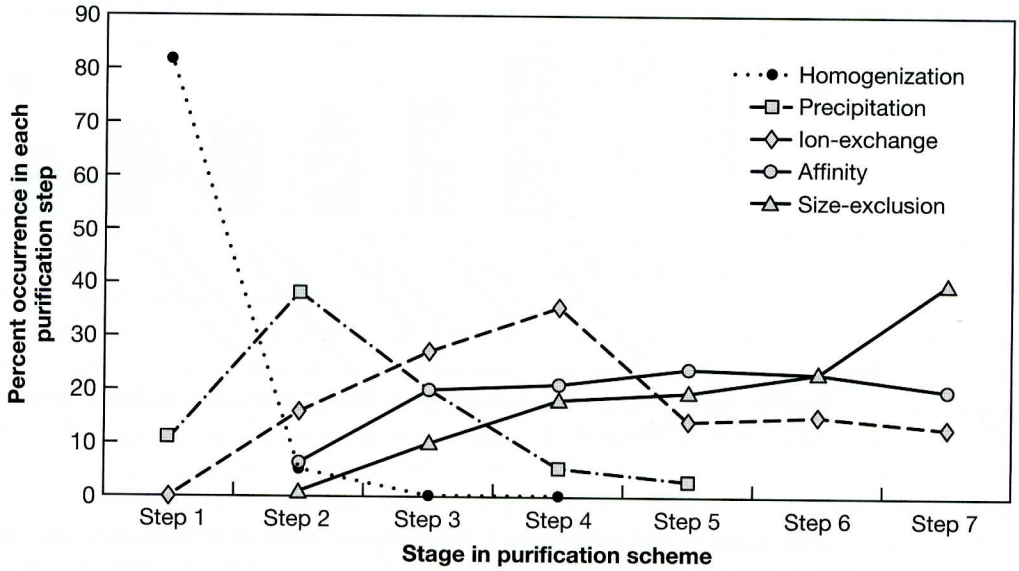


FIGURE 1.5. Order of purification steps in a generalized protein purification. (Adapted, with permission, from Bonnerjea et al. 1986.)

1.3). Although SEC has very low capacity for loaded protein, it serves an important role in removing self-aggregates of an otherwise homogeneous target protein (a necessary requirement for crystallization studies, NMR analysis, and physicochemical characterization of higher-order complexes). This step is also suitable for transfer of the protein to a volatile buffer if the protein is to be lyophilized for long-term storage.

If the target protein (especially low-molecular-weight proteins) is to be used for sequence determination and/or identification by peptide mapping of posttranslational modifications, RP-HPLC is the preferred final polishing step because of its efficiency in removing modified forms of target protein, such as proteolytic truncations and heterogeneous carbohydrate adducts. Proteins and peptides recovered from RP-HPLC columns, typically in small volumes of trifluoroacetic acid/acetonitrile, can usually be stored for long periods at –20°C (see Chapter 7; see also Chapter 5 in Simpson 2003).

Which Order of Steps Is Best?

Although one would not a priori expect the sequence of steps in a purification protocol to be a major consideration, in practice, this is the case. According to an analysis of 100 successful purification methods by Bonnerjea and co-workers (1986), homogenization is generally followed by clarification/fractional precipitation, then anion-exchange chromatography, affinity separation, and finally, SEC (see Fig. 1.5). Although new methods enhance final product purity, they are typically used in addition to established procedures and not in place of them. An important consideration in designing the order of purification steps is to select, where possible, a sequence that minimizes buffer-exchange steps. For example, hydrophobic interaction chromatography (where samples are applied to the column under high salt concentrations) can follow fractional precipitation (using ammonium sulfate) or ion-exchange chromatography (where proteins are eluted with high salt concentration) without the need for a buffer-exchange step (typically accomplished by SEC, dialysis, or membrane ultrafiltration). In contrast, the use of RP-HPLC for purifying native proteins is best suited to a late stage of purification, by which time extraneous compounds that would otherwise destroy the chromatographic support have been removed.

CHECKLIST FOR PROTEIN PURIFICATION

- *Define end goals.* Decide on the level of purity and quantity required for final target protein (see Table 1.1).
- *Establish a rapid analytical assay* to monitor the purification of the target protein. Fast detection of protein activity and recovery at each stage is essential for an efficient purification procedure. The assay should also be capable of easily handling a large number of samples.
- *In pilot experiments, define the chemical and physical characteristics* of the target protein (e.g., pI, size, temperature stability, and ligand specificity) in order to simplify the separation technique selection and optimization. Where possible, use a different separation technique at each purification stage.
- *Keep the purification procedure AS SIMPLE AS POSSIBLE:* Extra steps invariably reduce the overall yield of the target protein and increase the process time.
- *Minimize sample handling at every stage* and avoid lengthy procedures that might result in reduced recovery and loss of biological activity.
- *Remove damaging contaminants,* particularly proteases, early in the purification procedure.
- *Be careful with the addition of stabilizing additives* (e.g., detergents and salts), because they may need to be removed in subsequent purification steps or they may interfere with assays.

STRATEGIES BASED ON ELECTROPHORESIS FOR SEPARATING PROTEINS

As we have seen, the reason that it is possible to separate one protein from a mixture of thousands of proteins is that a number of their physical and chemical properties vary tremendously, especially their molecular size (or weight) and charge. The latter property, overall charge, results from proteins having different numbers and sequences of amino acids, especially those that have ionizable side groups. For example, the side-chain carboxyl moieties of the acidic amino acids, aspartic acid and glutamic acid, are negatively charged at pH values greater than their pK_a values (4.4–4.6) and uncharged (un-ionized) at pH values lower than their pK_a values (see Fig. 1.6). In contrast, the side chains of the basic amino acids, histidine (imidazole group), lysine (ε-amino group), and arginine (guanidine group), are positively charged at pH values lower than their pK_a values and are uncharged at pH values higher than their pK_a values (Fig. 1.6). Because these ionization reac-

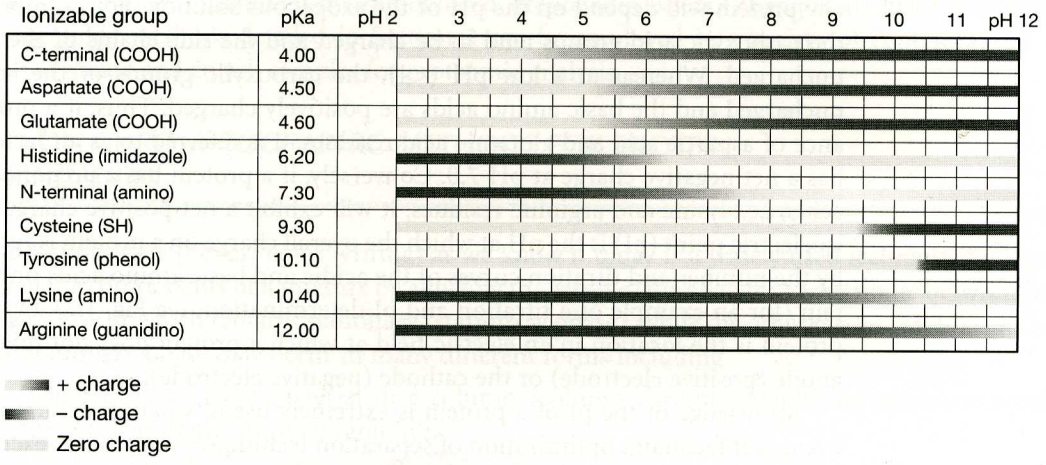


FIGURE 1.6. Charge of the ionizable groups found on native proteins as a function of pH.

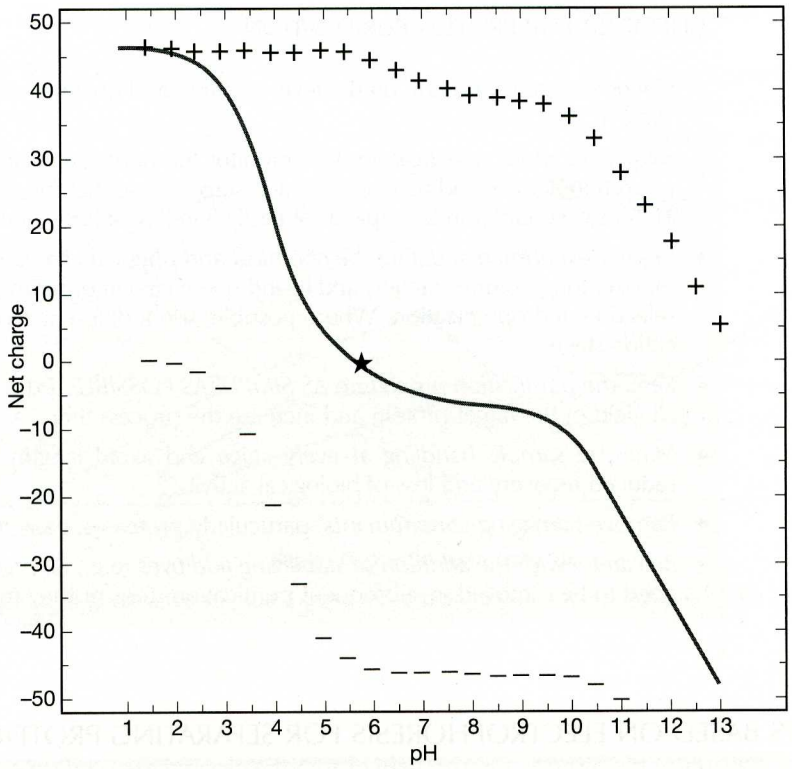


FIGURE 1.7. Titration curve and isoelectric point (pI) of *Escherichia coli* RNA polymerase transcription factor σ^{32}. Shown are theoretical plots of the number of positively charged and negatively charged amino acid side chains as a function of pH for the *E. coli* RNA polymerase transcription factor σ^{32}, based on its amino acid sequence. The pI of *E. coli* σ^{32}, indicated by the star, is 5.78. The number of each type of charged amino acids in the molecule is 23 arginines, 16 lysines, 6 histidines, 7 tyrosines, 22 glutamic acids, and 23 aspartic acids. The plot was generated using the Genetics Computer Group Sequence Analysis Software package. The usefulness of such a plot can be illustrated as follows. At pH 7.9, σ^{32} has a negative charge of 46 and a positive charge of 40, giving the molecule a net charge of –6. From this information, it is evident that σ^{32} is able to bind tightly to both anion- and cation-exchange columns, because its charge residues are not evenly distributed on the surface of the protein. Such behavior can be exploited to purify the protein, because most proteins will not bind to both types of ion exchangers under a single solvent condition. (Reproduced, with permission, from Marshak et al. 1996.)

tions are readily reversible, the proportion of charged and uncharged amino acid side groups in any protein will depend on the pH of the exogenous solution. For example, at a high pH (~12), the carboxylic acid groups tend to be charged and the side chains of the basic amino acids are uncharged. Whereas at a low pH (<4), the carboxylic groups of the acidic amino acids are uncharged and the basic amino acids are positively charged. Thus, if a protein has a preponderance of aspartic acid and glutamic acid residues, it is referred to as an "acidic protein" because it has a net negative charge at pH 7.0. Conversely, if a protein has a preponderance of basic amino acids, i.e., lysine and arginine residues, it will exhibit a net positive charge at pH 7.0. A protein's isoelectric point (pI) is the pH at which the overall charge on a protein is zero, and it is determined by the number and titration curves of the acidic and basic amino acids on the surface of the protein (for an example of a titration and pI determination, see Fig. 1.7). Operationally, the pI of a protein is the location in an electric field at which a protein does not migrate toward either the anode (positive electrode) or the cathode (negative electrode).

Knowledge of the pI of a protein is extremely useful when designing a purification strategy, because it facilitates optimization of separation techniques such as ion-exchange chromatography (Chapter 5), chromatofocusing, and the electrophoretic separation methods outlined in Section 1 (see Chapters 3 and 4).

Enrichment of Low-Abundance Proteins by Preparative Electrophoresis

Given the dynamic range of protein abundances in biological samples such as cell lysates and blood, there is a compelling argument for subfractionation of complex protein mixtures to study low-abundance proteins. For example, the dynamic range of protein abundance in a biological sample can be as high as 10^6, with protein abundances ranging from 10 copies per cell for transcription factors up to 1,000,000 copies per cell for the cytoskeletal proteins that maintain cellular architecture.

Traditionally, 2D gel electrophoresis (Klose 1975; O'Farrell 1975) has been the preferred proteomics technique for separating hundreds to thousands of proteins in a single experiment (for reviews, see Görg et al. 2000; Rabilloud 2002; Simpson 2003). Moreover, when narrow immobilized pH gradients and extended separation distances are employed, resolution can be further improved (Wildgruber et al. 2000). Because 2D gel electrophoresis can separate only a subset of a total proteome, at best 1500–2000 proteins, this method is limited to the most abundant proteins (Gygi et al. 2000). In the small sample volumes (~10–300 µl) typically used for proteomic analysis by 2D gel electrophoresis, a large percentage of the expressed proteins are not present in sufficient amounts to be detected by current mass spectrometric methods (Gygi and Aebersold 2000; Simpson 2003). It is also difficult to resolve proteins of very high molecular weight, very low molecular weight, and very basic or very hydrophobic proteins, such as membrane-associated proteins, using this method. However, recent progress has been made in adapting 2D gel electrophoresis to high- and low-molecular-weight proteins (Tastet et al. 2003) and the judicious use of nonionic and zwitterionic detergents to solubilize membrane proteins (Luche et al. 2003; Rabilloud 2003). For the enrichment of low-abundance proteins from crude cell extracts, several protein-enriching methods have been described for separating the original protein mixture into simpler fractions each containing a lower number of proteins than the starting material. One approach takes advantage of the macromolecular architecture of a cell—the subcellular compartments, organelles, macromolecular structures, and multiprotein complexes—to break down the problem into "bite-size" components. This has led to the description of subcellular proteomes and "subproteome" analysis (for a review, see Jung et al. 2000 and articles reported in a special issue of the journal *Proteomics* [Huber 2003]). Another approach for protein enrichment involves differential detergent fractionation using detergents such as Triton X-114 (Chapter 2). Enrichment of proteins from larger volumes can also be performed by preparative polyacrylamide gel electrophoresis on the basis of protein size, usually in the presence of ionic detergents such as SDS and a preparative SDS-PAGE apparatus such as the Bio-Rad Model 491 PrepCell (Zugaro et al. 1998; Fountoulakis and Juranville 2003).

Other prefractionation methods employed to enrich fractions prior to 2D gel electrophoresis include selective precipitation, for example, with trichloroacetic acid/acetone (Görg et al. 1998), and affinity purification to isolate plasma membranes, for example, using cell surface biotinylation with affinity enrichment by immobilized streptavidin beads (Zhang et al. 2003). The chromatographic methods detailed in Section 2 can be employed in a variety of configurations to prefractionate and enrich samples for proteomic analysis.

STRATEGIES BASED ON CHROMATOGRAPHIC METHODS FOR PROTEIN AND PEPTIDE PURIFICATION

Chromatography (Greek: color writing) as we know it today is a widely used technique for separating the components of a mixture by allowing the sample (the analyte) to distribute between two phases, one of which remains stationary (stationary phase), while the other moves (mobile phase). The stationary phase may occur in many different forms including

- a packed bed of solid material in a column (column chromatography, or more commonly referred to as liquid chromatography),
- spread as a thin layer or film on a flat plate (thin-layer chromatography), or
- paper (paper chromatography).

The mobile phase may be gaseous (as in gas chromatography) or liquid. Only a liquid mobile phase is suitable for the separation of proteins and peptides. Whenever the individual components in the analyte distribute between the stationary and mobile phases to different extents, separation will occur. Compounds with a greater tendency to stay in the stationary phase will migrate through the system at a slower rate than those that favor the mobile phase.

One of the most common methods for separating proteins and peptides is column chromatography, also called liquid chromatography (LC). The stationary phase is packed into a tube or column, made of metal, plastic, or glass, through which the liquid mobile phase (the eluent, which is most often an aqueous buffer) is passed. Separation of analytes is based on their dynamic binding interaction or partitioning between the mobile phase and the surface of the stationary phase. By manipulating the chemistry of the stationary phase and/or composition of the mobile phase, along with other operating conditions (e.g., the gradient shape of eluent solvent), it is possible to obtain very subtle separations due to minor differences in molecular structure (e.g., a single amino acid modification such as an oxidized methionine or an amino acid substitution) or protein configuration (e.g., native versus unfolded structure).

Several different modes of LC allow proteins and peptides to be separated on the basis of their relative size, shape, charge, hydrophobicity, and bioaffinity. These different versions of LC are characterized by different forms of stationary phase (for a summary, see Table 1.5).

As depicted in Figure 1.8, chromatographic separation systems require the following key elements:

- ***A stationary phase*** with controlled structure (particle size, shape, etc.) and surface chemistry (bonded phase).
- ***A column*** packed with stationary phase.
- ***A mobile phase(s)*** or solvent(s) of controlled chemical composition that moves the solute through the column. For most chromatographic separations, the notable exception being size-exclusion chromatography, columns are developed using either two or three solvents that are blended in a controlled manner to produce an accurate and reproducible gradient.
- ***Chromatography equipment*** capable of accurately delivering the sample (sample injection) and controlling the blend (mobile-phase composition) and the flow of the mobile phase(s) to the column, as well as detecting the fractionated analytes.
- ***Software programs*** for blending the mobile phase(s) and running them through the column to effect the separation (i.e., procedures for developing the column).

Stationary Phase

Stationary phases (also referred to as column packings, gels, or media) are generally considered to be the key element of LC. Each is designed to provide a large surface area that is accessible to the

TABLE 1.5. Liquid chromatography methods for separating proteins and peptides

Principle of Chromatographic Separation	Type of Chromatography
Net charge	Ion-exchange chromatography (Chapter 5)
Size and shape	Size-exclusion chromatography, also referred to as gel-filtration or gel-permeation chromatography (Chapter 6)
Hydrophobicity	Hydrophobic interaction chromatography
	Reversed-phase high-performance liquid chromatography (RP-HPLC) (Chapter 7)
Biological function (bioaffinity)	Affinity chromatography (Chapter 8)
Carbohydrate content	Lectin chromatography (Chapter 8)
Antigenicity	Immunoaffinity chromatography (Chapter 8)
Metal binding	Immobilized metal ion affinity chromatography (Chapter 9)

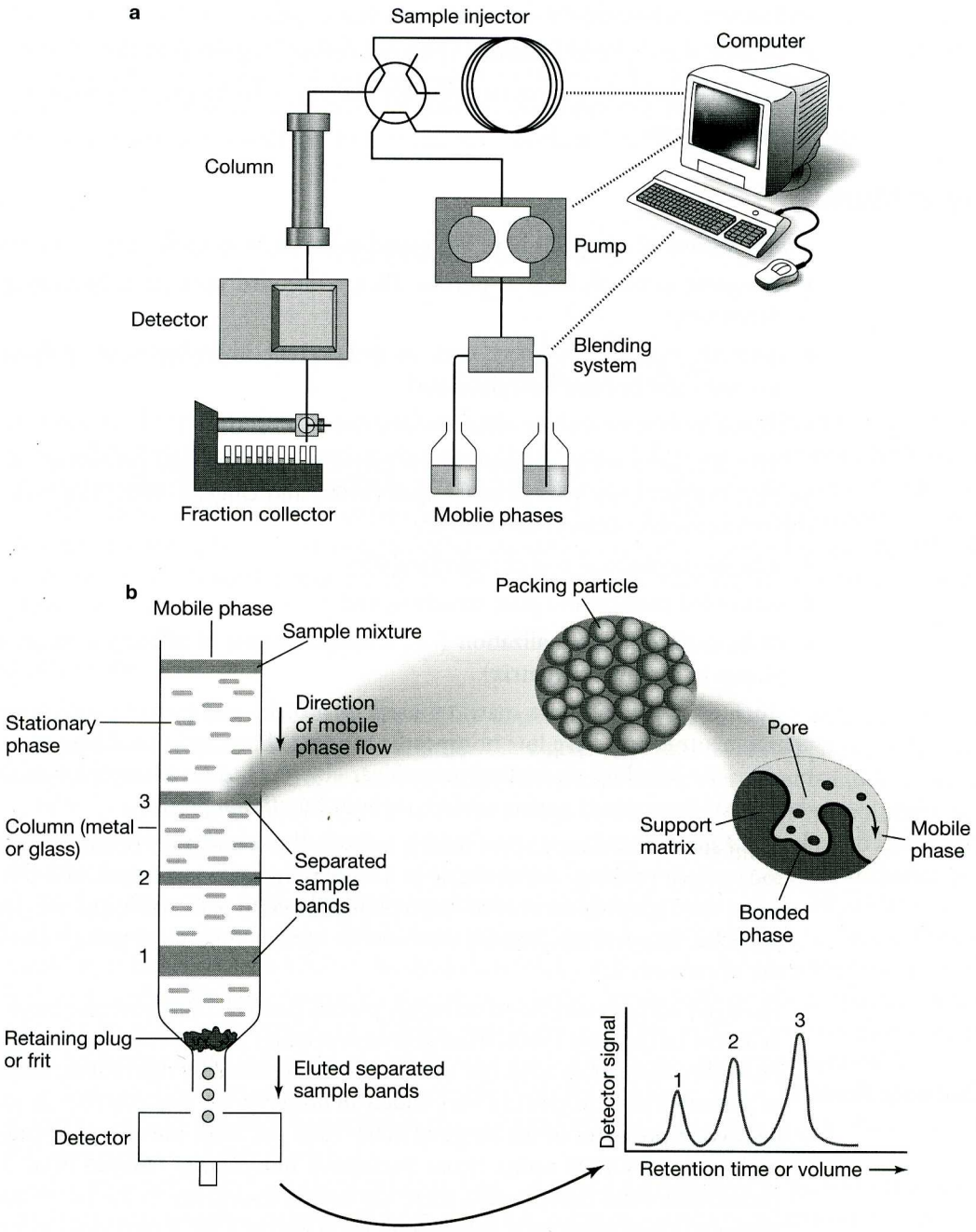


FIGURE 1.8. Basic elements of liquid chromatography. (*a*) Components of an LC system and their connections. (*b*) Details of an LC column and its packing. (Adapted, with permission, from PerSeptives Biosystems, Inc. 1996.)

mobile phase and sample molecules. This is usually accomplished by using small, highly porous particles, termed the support matrix, in which all or most of the surface area within the pores is accessible to the mobile phase. These particles form the backbone or "skeleton" of the stationary phase and provide a surface that can be chemically coated with a bonded phase containing functional groups that provide the desired specific binding interaction. Ideally, both the bonded phase and support matrix exhibit minimal nonspecific binding interactions with the analyte. Column packings for LC are classified according to the following features:

- ***Support matrix:*** rigid solids, hard gels, or soft gels.
- ***Particle size (d_p) and structure:*** spherical versus irregular particles.
- ***Pore structure:*** porous versus pellicular and superficially porous particles.
- ***Bonded phase.***

Support Matrix

A wide variety of materials have been used as chromatographic support matrices, including

- inorganic materials such as porous silica, controlled pore glass, hydroxyapatite, alumina, and zirconium;
- synthetic organic polymers such as polystyrene-divinylbenzene, polyacrylamide, polyvinyl alcohols, and polymethacrylate; and
- natural polymers such as cellulose, dextran, and agarose.

For detailed reviews describing the physical and chemical properties of various chromatographic supports, see Mikes and Coupek (1990) and Unger (1990). The basic requirements of all chromatographic support matrices are

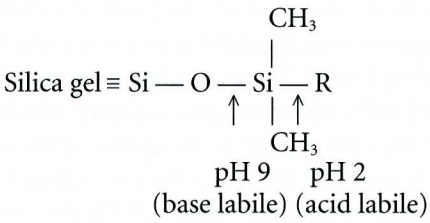
- suitable mechanical and chemical stability,
- controlled particle and pore structure, and
- facile surface functionalization (i.e., a simple means of affixing a range of different bonded phases to the surface matrix).

In most cases, support matrices used in protein and peptide applications are hydrophilic, charge-neutral, and have low nonspecific binding characteristics. These are especially important because the bonding-phase treatment invariably fails to go to completion, leaving a small percentage of the support matrix surface without functional groups.

- ***Rigid solids*** based on a silica matrix are still the foundation of most column packings in use today. Such packings are available in a wide range of sizes, shapes, and porosity and can withstand the high pressures (4000–6000 psi) required to pack stable and efficient columns of small particles. Importantly, bonded phases with various functional groups can be readily affixed to the silica surface.
- ***Hard gels*** are generally based on highly porous particles of polystyrene cross-linked with divinylbenzene. In the early 1960s, these gels were popular for ion-exchange chromatography and size-exclusion chromatography, but were gradually replaced by rigid solids. Depending on the mode of preparation, hard gels can vary widely in both rigidity and porosity. In practice, hard gels can be used at pressures in the range of 2000–5000 psi. With the advent of high-speed chromatography using POROS resins (from PerSeptive Biosystems) (Afeyan et al. 1990) and SOURCE (from Amersham Biosciences), there has been a resurgence in the use of hard gels. Unlike conventional chromatography particles, POROS and SOURCE particles have two distinct types of pores: large pores that transect the particle and short diffusive pores that branch off from the throughpores. Compared with conventional chromatographic packings used under optimal chromatographic conditions, this design enables chromatographic separations to be carried out much faster, with little or no loss in resolution or capacity (Moritz and Simpson 1995).
- ***Soft gels*** (such as cellulose, dextran, polyamide, and other hydrophilic polymers), in contrast to the rigid gels and hard gels, cannot withstand high pressures. Soft gels are widely used for the separation of proteins based on size and shape (see Chapter 6).

It is useful to examine the structure and properties of two major types of chromatographic supports: silica and organic polymer gels. Both are manufactured as spherical beads of various sizes, often with extremely narrow size distributions. Both types of packings have been synthesized with gradations in pore size, and also as nonporous particles. The chemical resistance, especially

pH stability, is a major issue regarding the utility of these supports. For instance, porous silica exhibits a somewhat pH-dependent solubility of ~100 ppm in the pH 2–8 range, but begins to dissolve rapidly in aqueous solutions above pH 9 and below pH 2. The bond between the oxygen and silicon is unstable in alkaline medium, and the bond between the silica and carbon atoms of the functional group R is unstable in a strong acid medium. This is shown schematically:



Although polystyrene-divinylbenzene-based gels are stable over a wider pH range (e.g., pH 2–12), they suffer from the slow release of monomer. In the case of dextran and agarose, both are stable toward alkaline media, but are attacked by strong acids. Both silica and the organic polymer gels exhibit matrix effects, due to parent surface-active groups that lead to nonspecific binding. Whereas the soft gels (also termed xerogels) can swell/shrink in the presence and absence of solvent, the rigid gels/hard gels (also referred to as aerogels) are rigid and their particle diameter is independent of solvent.

Particle Size and Structure

Particle size (d_p). The particle size is a critical determinant in LC, because it influences the chromatographic efficiency (and hence resolution) in a given separation. To a great extent, the distribution of particle shape and size also determines the permeability of a column and its mechanical stability (i.e., the column lifetime). It is the surface area of the bonded stationary phase and its accessibility to the mobile phase that controls analyte retention, which is reflected in the analyte capacity factor k' and the loadability of the column (see Chromatographic Performance, p. 23).

Particle shape. It is preferable to work with particles having a narrow range of particle sizes (i.e., with d_p varying no more than 1.5-fold from the smallest to largest particles), because column permeability is largely determined by the smallest particles in the column, whereas column efficiency is determined by the larger particles. Note that packed columns with very broad particle distribution are typically both inefficient and less permeable, whereas well-packed columns containing particles having a narrow range of particle sizes are both efficient and more permeable.

Chromatographic media may be either spherical or irregular in shape. Although spherical and irregular particles can each be packed to give columns of equal chromatographic efficiency, spherical materials are greatly preferred.

Recent advances in packing technology have led to the development of porous microparticles (3–20 μm in diameter d_p) that yield high-efficiency columns with fast separation times and moderate capacities and, more recently, to high-speed perfusion chromatography particles, such as the POROS and SOURCE particles. Typical particle sizes used for various LC applications are presented in Table 1.6.

TABLE 1.6. Typical particle sizes used for various LC applications

Purpose of LC	Particle Size
Analytical applications	3–10 μm diameter[a]
Preparative separations	10–40 μm diameter
Low-pressure/large-scale applications	40–150 μm diameter
Very large-scale operations	~300 μm diameter

[a]Media made from particles with a d_p <1–3 μm have proven to be impractical because of inherent problems with packing and the need for very high operating pressures.

Pore Structure (Accessible Surface Area)

Most LC packings are designed with as large a surface area as possible to provide easy access of the mobile phase and sample molecules to the bonded phase. This is accomplished using a porous resin of small particle size and with pores whose diameters approach the molecular sizes of the solutes. Typically, the pore diameter must be ~5 times the size of the molecules being purified to permit them to access all of the pores via molecular diffusion. Thus, selection of a chromatographic support with the desired porosity permits the free diffusion of analytes into and out of the pores, enabling them to optimally interact with the stationary phase. The pore structure of the packing critically affects column capacity, which is the amount of sample material that can either bind to or be separated by a chromatographic column. Since the surface area per unit volume of a particle is inversely related to the pore diameter, the use of particles with overly large pores will result in a loss of capacity—hence, both the average pore diameter and pore size distribution are critical chromatographic parameters. Chromatographic packings can be grouped according to their pore dimensions:

- Macroporous packings contain pores ranging from 1000 to 10,000 Å.
- Mesoporous packings, also referred to as "wide-pore" packings, contain pore diameters that are intermediate between the micro and macroporous resins (e.g., 180–500 Å).
- Microporous packings have 60–120-Å pore diameters.

Size-exclusion chromatography media are designed with a range of different pore diameters, which permits the separation of molecules of different sizes (see Chapter 6). Reversed-phase media, such as silica, with pore sizes of ~60 Å and 120 Å are used predominantly for small peptides, whereas the separation of large proteins—by RP-HPLC, ion-exchange chromatography, or hydrophobic interaction chromatography—requires pore diameters of >300 Å.

Bonded Phase

Whereas the column packing matrix provides the chemically inert "skeleton" for the stationary phase, the bonded phase provides the functional groups, which are designed to selectively bind solute molecules, thus separating them from the other components in the sample mixture. To minimize interference with the functional groups, the underlying packing matrix should be "neutral" and have minimal nonspecific binding with the sample.

The bonded phases of silica packings are prepared by reacting the surface silanol groups with an appropriate chlorosilane (both monofunctional and bifunctional silanes can be used for this purpose) (Fig. 1.9A). For LC particles that have polystyrene as a rigid matrix (e.g., POROS media), the bonded-phase chemistry involves first adsorbing and then cross-linking a copolymer (containing both hydrophobic and hydroxyl groups) onto the surface of the polystyrene. Once cross-linked, the hydroxyl groups are then functionalized to form the final bonded phase (Fig. 1.9B). The functional R groups are usually methyl (CH_3) groups, whereas the nature of the R′ group can be varied to give both apolar and polar stationary phases. Typically, R′ is a hydrocarbon chain (C_6, C_8, or C_{18}) that gives rise to an apolar phase (see RP-HPLC in Chapter 7). However, polarity can be introduced by substituting the terminal methyl ($-CH_3$) group in the hydrocarbon chain, for example, by a nitrile group (-C N) or an amino group ($-NH_2$). For ion-exchange chromatography, the functional groups can be either acidic (cation exchangers) or basic (anion exchangers). Cation exchangers contain acidic groups that are referred to as weak (e.g., $-COO^-$) or strong (e.g., $-SO_3^-$), whereas anion exchangers contain basic groups that may be weak (e.g., $-NH_2$) or strong (e.g., $-NR_3^+$).

Several characteristics of the bonded phase critically influence chromatographic behavior, especially the selectivity and capacity of the column. Key among them are the precise chemical structure of the functional groups, the manner in which the functional groups are chemically bonded to the surface matrix surface, and the bonding density of the functional groups. Other

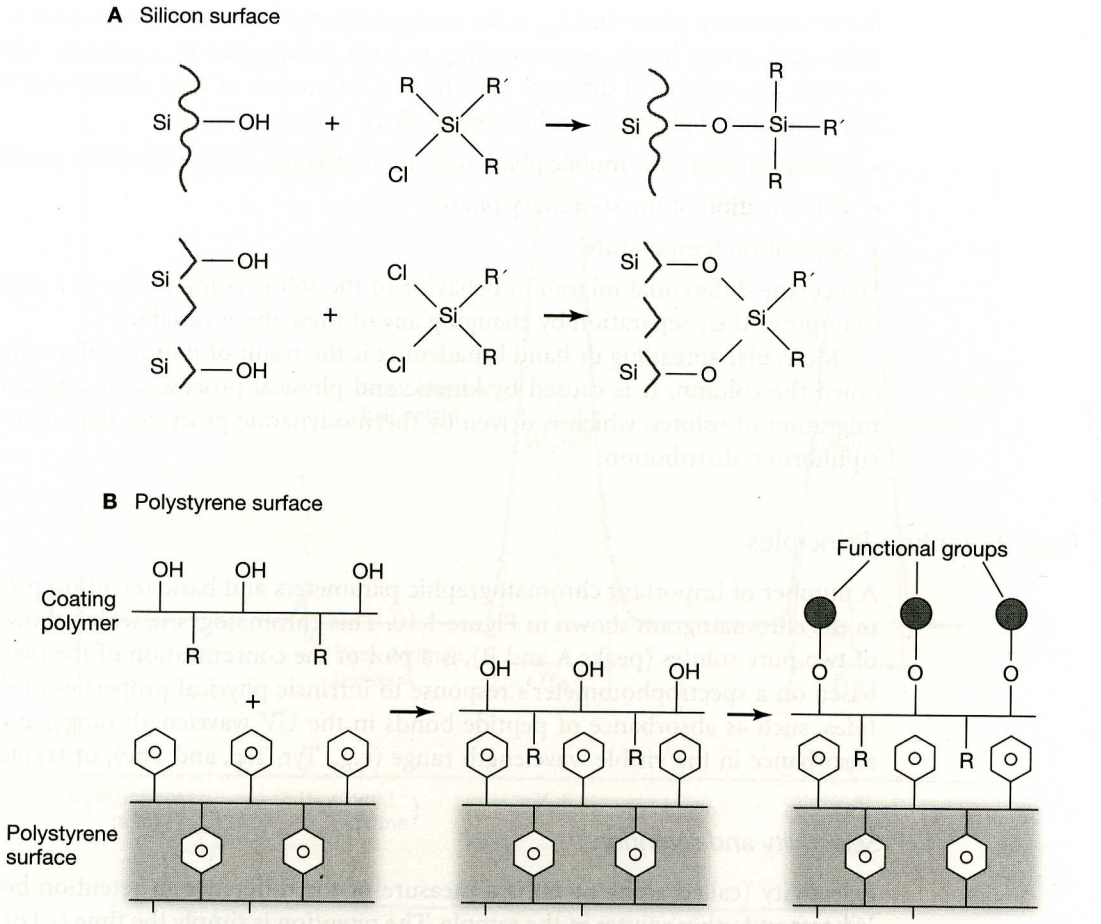


FIGURE 1.9. Bonded-phase chemistries used to affix functional groups to inert stationary-phase particles. (*A*) Solid silica particle surface. Reaction of silanol groups with monofunctional (*top*) and bifunctional silane (*lower*). (*B*) Rigid polystyrene-divinylbenzene particle surface. Bonded-phase chemistry system used with POROS media. (Redrawn, with permission, from PerSeptives Biosystems, Inc. 1996.)

important considerations are the chemical and physical stability of the bonded phase under normal operating, storage, and regeneration conditions.

CHROMATOGRAPHIC PERFORMANCE

The separation of charged molecules, such as peptides and proteins, as they move down a column is affected by (1) the differential migration of solutes and (2) their spreading or dispersion (also referred to as peak or band broadening).

Differential migration refers to the variable flow of solutes as they move down a column. Each solute partitions in a unique way between the stationary phase and the mobile phase. It is these different equilibrium distributions that cause the solutes to migrate through the column at different rates. When the interaction of a solute with the stationary phase is very strong, it is retained to a greater extent, and thus will move through the column more slowly than another solute in the sample that interacts less strongly. Such behavior can be defined by the equilibrium distribution coefficient (or partition coefficient, K_D) as $K_D = S_S/S_M$, where S_S is the concentration of a solute (S)

in the stationary phase and S_M is the concentration of the same component in the mobile phase. Individual solute bands corresponding to each component in a mixture will therefore migrate through the column at different velocities as a function of their differential migration behavior. This migration behavior is influenced by three major variables:

- Composition of the mobile phase (e.g., ionic strength, pH, and organic modifier concentration).
- Composition of the stationary phase.
- Separation temperature.

Hence, the differential migration behavior of the solute components in a mixture can be altered to improve their separation by changing any of these three variables.

Molecular spreading or band broadening is the result of dilution of a solute band as it moves down the column. It is caused by kinetic and physical processes, in contrast to the differential migration of solutes, which is driven by thermodynamic processes that arise from differences in equilibrium distribution.

Basic Retention Principles

A number of important chromatographic parameters and basic retention principles are depicted in the chromatogram shown in Figure 1.10. This chromatogram, which shows the elution profile of two pure solutes (peaks A and B), is a plot of the concentration of the two solutes. The plot is based on a spectrophotometer's response to intrinsic physical properties of the proteins or peptides, such as absorbance of peptide bonds in the UV wavelength range, amino acid side-chain absorbance in the visible wavelength range (e.g., Tyr, Trp, and Phe), or tryptophan fluorescence.

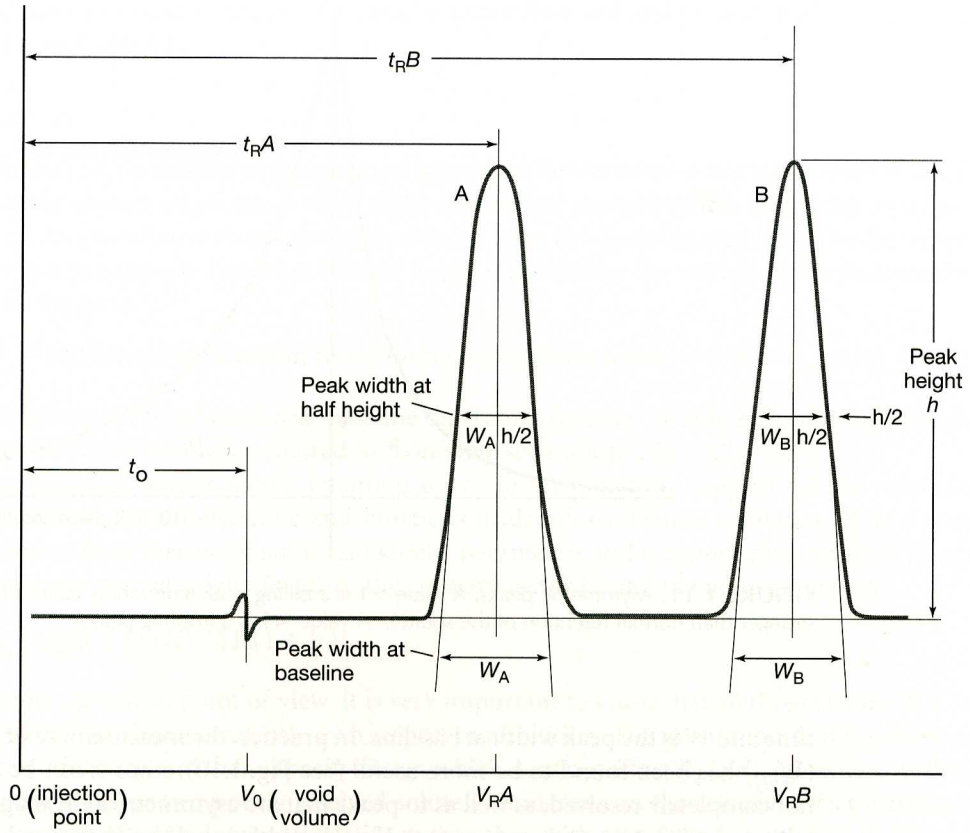
Selectivity and retention

Selectivity (called alpha or α) is a measure of the difference in retention between the solute of interest and other solutes in the sample. The retention is simply the time (t_r) or volume (v_r) it takes for a solute to move through the column, from the point of injection to the spectrophotometer, where it is monitored as an eluted peak. If a solute does not interact with the stationary phase, it will pass directly through the column and elute at time (t_o) in the void volume (v_o). The void volume is a column characteristic that represents both the interstitial volume between the particles of the stationary phase and the available volume within the particle pores. For two solutes to be resolved, they must exhibit different equilibrium distribution coefficients such that they are retained for distinct periods of time (i.e., different t_r values) and elute in different solvent volumes (v_r). The capacity factor (k')—also referred to as the retention factor—is normalized retention under isocratic elution conditions. It is a unitless term that is a measure of the retention behavior (i.e., the degree of retention) for a particular solute on a particular column, assuming equal flow rates. The capacity factor k' can be calculated by the following equation:

$$k' = (t_r - t_o)/t_o = (v_r - v_o)/v_o$$

where k' is the number of column volumes required to elute a particular solute, and t_o and v_o represent the void time and void volume, respectively. Thus, k' is directly related to the distribution coefficient (or partition coefficient) of a solute between the mobile and stationary phases (i.e., moles of solute in stationary phase per moles of solute in mobile phase) and is now well understood in both empirical and thermodynamic terms.

Selectivity is sometimes expressed as the ratio of the capacity factors k' of two solutes being separated: $\alpha = k'_2/k'_1$. In gradient elution, k' is not a valid measurement, and the simple retention time (t_r) or retention volume (v_r) is used. Selectivity is affected by the surface chemistry of the column packing, the nature and composition of the mobile phase, the nature of the stationary phase, and the gradient shape.



Retention	$k' = (t_R - t_o) / t_o = (V_R - V_o) / V_o$
Selectivity	$\alpha = k'_B / k'_A$
Efficiency	$N = 16\,(t_R / W)^2 = 5.54\,(t_R / W_{h/2})^2$
Resolution	$R_s = (t_RB - t_RA) / (W_A + W_B)^{1/2}$
	$= (t_RB - t_RA) / (W_{Ah/2} + W_{Bh/2})0.85$
	$= 1/4\,(\alpha - 1)\,(N^{1/2})\,(k'/1 + k')$

FIGURE 1.10. Important chromatographic terms and how they are measured.

Band broadening and efficiency

As solute peaks or zones migrate down a column or through a chromatography system, they usually increase in volume. This band broadening (or peak spreading) phenomenon that the solute experiences during transit through the column is an unavoidable consequence of LC, and keeping it to a minimum is a technical challenge for the researcher. The extent of band broadening that occurs during elution is reflected in the column efficiency, which is traditionally expressed in terms of the theoretical plate number N of the column: $N = 16\,(t_r/W)^2$, where t_r is the retention

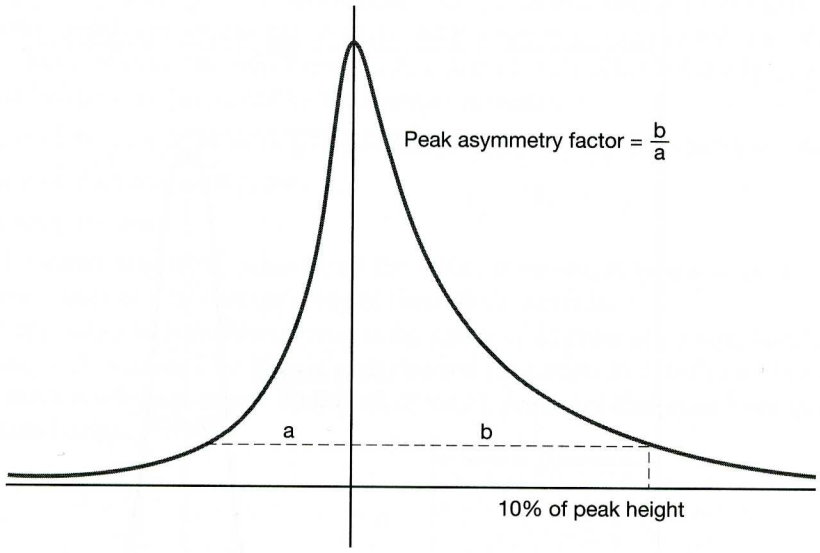


FIGURE 1.11. Asymmetric peaks. A value >1 is a tailing peak (commonly caused by sites on the packing with a stronger than normal retention of the solute). A value <1 is a fronting peak.

time and W is the peak width at baseline. In practice, the measurement of peak width at half height ($W_{h/2}$) has been found to be more useful (see Fig. 1.10), since it can be applied to peaks that are not completely resolved, as well as to peaks that are asymmetrical in shape (e.g., peaks that exhibit tailing; see Fig. 1.11 and asymmetry). If peak width at half height is used, then $N = 5.54\ (t_r/W_{h/2})^2$.

The value N is a useful measure of column performance. For a given set of operating conditions (i.e., for a particular column, mobile phase, fixed mobile-phase velocity, and operating temperature), N is approximately constant for different solute bands in a chromatogram. Hence, N is a measure of column efficiency, and, in general, the higher the number of theoretical plates, the better the column. In other words, a column with a high N value will provide solute peaks with narrow bandwidths (small W values) and improved separations. Although the value for N is largely independent of t_r and remains constant for different solute bands in a chromatogram, the solute peak widths increase proportionately with t_r, so that later-eluting solute bands typically show a decrease in peak height and eventually disappear into the baseline.

Since the quantity N is proportional to column length L, an increase in L usually results in an increase in N and hence a superior separation. This is especially so for LC operated under isocratic (i.e., constant-composition mobile phase) elution conditions. In gradient elution chromatography, all solute bands in the chromatogram tend to be of uniform width. The proportionality of N and L can be expressed as follows

$$N = L/H$$
$$H = HETP = L/N$$

where H is the height equivalent of a theoretical plate (plate height) or *HETP* and L is the length (usually in mm) of the column. Hence, *HETP* is a better measure of column efficiency than N because it permits a better comparison between columns of different lengths that are operated under identical chromatographic conditions. Thus, small values of H are indicative of more efficient columns. For a well-packed HPLC column of 5–μm particles, *HETP* or H values usually range from 0.01 to 0.03 mm.

Band broadening can also occur in parts of the chromatographic system other than the column, especially when there is excessive dead volume within the system. These extra-column contributions to band broadening can be minimized through careful attention to the type of injector

used, choice of tubing, length of tubing, detector flow cell design, and dead volume of fittings, unions, and adaptors.

Resolution

Resolution (R_s) is defined as the extent of separation between two chromatographic peaks. Clearly, it is of the utmost importance, since without adequate resolution, it is impossible to achieve separation. In quantitative terms, the resolution between two peaks (*A* and *B*) can be described as the difference in retention (in either volume or time) divided by the average of the peak widths at the base of the peak.

$$R_s = (\text{difference in retention time})/(\text{average peak base width}) = 2\,(t_{rB} - t_{rA})/(W_A + W_B)$$

A resolution of 1.0 indicates near baseline separation (overlap of only 2%), although a higher resolution (>1.5) is normally required for complete separation (Fig. 1.12).

As discussed earlier, solute retention on a chromatographic support can be rationalized in terms of both thermodynamic and kinetic considerations. Hence, resolution R_s is a composite function of both thermodynamic and kinetic parameters and is expressed in terms of an equation that includes the selectivity factor α, the capacity factor k', and the plate number N.

$$R_s = {}^1/_4(\alpha - 1)\,(N)^{1/2}\,[1/(1 + k')]$$

From a practical point of view, it is very important to know that all three factors, σ, k', and N, can be optimized independently with respect to resolution. The most important is selectivity, then column efficiency, and then capacity factor, which does not affect R_s to a large extent. As illustrated in Figure 1.13, for two poorly resolved solutes, an increase in the selectivity factor α results in a displacement of one solute peak relative to the other, with a profound increase in R_s. This increase in selectivity can be achieved by changing the mobile-phase pH, varying the mobile-phase solvents, altering the composition of the stationary phase and parameters such as the carbon load, or adjusting the temperature at which the separation is run.

An increase in efficiency (i.e., plate number N) causes a narrowing of the two solute peaks and a concomitant increase in peak height without affecting the retention times. Improving R_s by increasing N can be achieved by decreasing particle size, by increasing the column length, or by decreasing the mobile-phase velocity (i.e., flow rate).

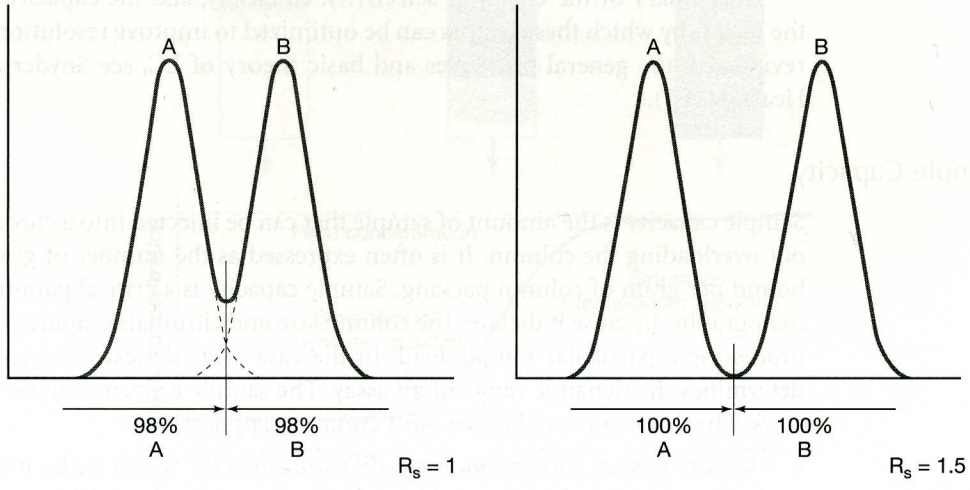


FIGURE 1.12. Separation results with different resolution. (Reprinted, with permission, from Amersham Biosciences 1999.)

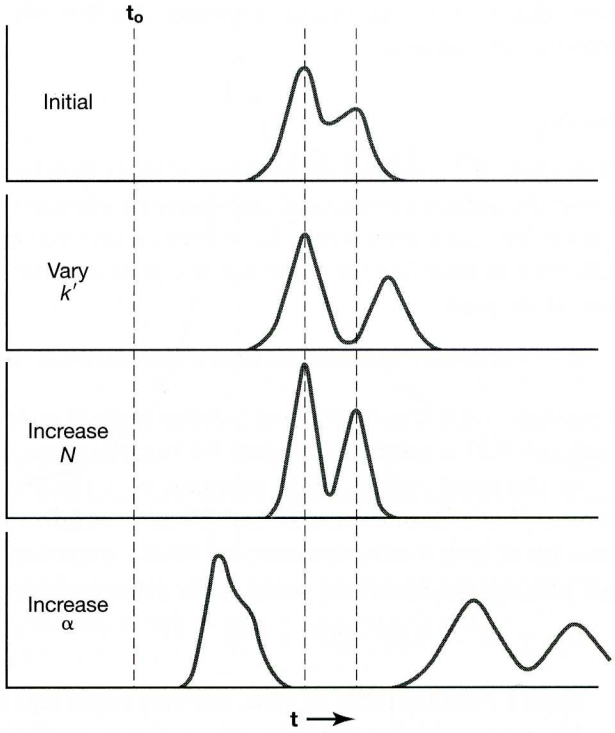


FIGURE 1.13. Effect of changes in k', N, or α on sample resolution R_s. (Reprinted, with permission, from Snyder and Kirkland 1979 [©Wiley].)

If k' for the initial separation is in the range $0.5 < k' < 2$, then a decrease in k' leads to a deterioration in the separation, whereas an increase in k' results in an improved separation. However, as k' is increased, peak heights tend to decrease with a concomitant increase in separation times (see Fig. 1.13). When k' is already within the optimal range, $1 \leq k' \leq 10$, but resolution must be improved; rather than changing the packing type, the best approach is to try increasing the N value by using a small-particle (≤10 μm) column, by increasing column length, or by decreasing the mobile-phase flow rate (at linear velocity, υ). With respect to capacity factor k' (or retention factor), resolution can be improved by changing the eluent strength.

A summary of the effects of selectivity, efficiency, and the capacity factor on resolution, and the means by which these factors can be optimized to improve resolution, is given in Table 1.7. For reviews on the general principles and basic theory of LC, see Snyder and Kirkland (1979) and Hearn (1991).

Sample Capacity

Sample capacity is the amount of sample that can be injected into a chromatographic system without overloading the column. It is often expressed as the number of grams of sample that can be bound per gram of column packing. Sample capacity is a critical parameter for preparative chromatography, because it dictates the column size and chromatographic system that are required for processing a particular sample load. In the case of analytical chromatography, sample capacity determines the dynamic range of an assay. The sample capacity can be measured in a number of ways. The following are the two most common applications:

- ***Measurement of saturation or equilibrium capacity,*** which is the maximal amount of protein that can be bound to the packing in a given mobile phase. This is accomplished by mixing a predetermined amount of packing with an excess of sample in a given mobile phase, allowing

TABLE 1.7. Effects of selectivity, efficiency, and the capacity factor on resolution

	R_S	=	$^1/_4(\alpha - 1)$	x	$N^{1/2}$	x	$k'/(1 + k')$
	Resolution	=	selectivity factor	x	efficiency factor	x	retention factor

Factor	Effect on R_S	How to Improve R_S
Selectivity factor ($\alpha = k_2/k_1$)	For closely spaced peaks, α is close to 1.0, so *small* changes in α have *large* effects on resulting resolution.	Alter composition of mobile phase (e.g., organic modifier pH, buffer salt), stationary phase, and/or temperature
Efficiency factor $N = 5.54\ (t_r/W_{h/2})^2$	Since R_S is a function of the square root of N, large changes in N are required to make small changes in resolution.	Increase column length, decrease particle size of column packing, or decrease flow rate. Minimize extra-column dead volume.
Retention factor ($k'=[t_r - t_o]/t_o$)	When k' is small (<1), R_S increases rapidly with an increase in k'. However, beyond a k' value of 5, R_S increases very little with further increases in k'. Separations that involve k' values >10 result in long separation times and excessive band broadening.	Alter the eluent strength. Values of k' can be increased and/or decreased by using so-called weaker and stronger solvents, respectively (see Table 7.2).

the mixture to come to equilibrium (typically, 16 hours), and then measuring the bound versus free binding molecule.

- ***Frontal adsorption analysis,*** which measures the capacity in a packed column under flowing conditions. The sample is loaded onto the column continuously until all of the binding sites on the packing are occupied (i.e., the column is completely saturated) and the sample concentration in the eluate equals the concentration being applied to the column (the latter is called the feed concentration).

Plotting the eluate concentration against the feed volume reveals the equilibrium (or saturation) capacity, which is the amount of sample that must be applied to the column to reach an eluate concentration equal to half the feed concentration (see Fig. 1.14). Another useful term,

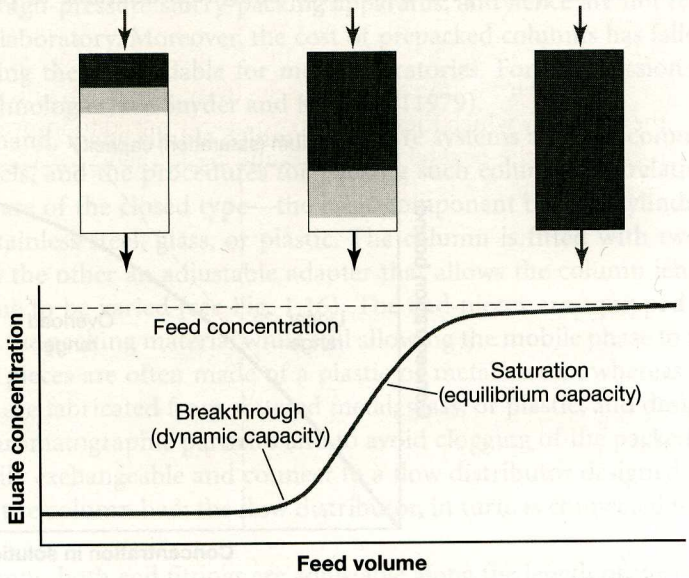


FIGURE 1.14. Frontal adsorption approach for determining the dynamic and equilibrium capacities of a column. (Adapted, with permission, from PerSeptives Biosystems, Inc. 1996.)

dynamic capacity, is the amount of sample injected onto a column in order to register the measurable breakthrough from the column. This value is typically equal to an eluate concentration of 1%, 5%, or 10% of the feed concentration (see Fig. 1.14).

To quantitate the amount of sample in a solution by analytical chromatography, it is crucial that the amount of bound sample be linearly dependent on the concentration of sample injected. This linear range is best determined by generating an adsorption isotherm, which establishes the relationship between bound and free concentration of solute. Figure 1.15 shows that at low concentrations of sample, a linear relationship is observed between bound sample and the concentration of sample applied to the column. As the application concentration increases, the bound concentration asymptotically approaches the saturation capacity, and the column is operating in an overload range. Although this overload range is deleterious for analytical chromatography, it is the normal operating range for preparative chromatography.

For optimal chromatographic performance and to achieve the greatest resolution, column loadability is a critical parameter. Loadability can be defined as the maximum amount of sample load that can be accommodated without affecting peak bandwidth. Table 1.8 illustrates this for columns of varying internal diameters. Under optimal chromatography conditions, in which peak volumes are at a minimum, the following protein load levels can typically be achieved (Simpson and Nice 1989):

2.5–5.0 µg for a 1.0-mm I.D. column
10–20 µg for a 2.1-mm I.D. column
30–100 µg for a 4.6-mm I.D. column

Chromatography Equipment

Packing a column

The optimal procedure for packing stationary-phase particles into a column is largely determined by the nature and size of the component particles. These parameters will vary depending on whether rigid solids, hard gels, or soft gels are being considered. The column components (e.g. tubing and fittings) that are used can also have a significant effect on the choice of packing method. It is important that tubing and fittings be chosen to avoid excessive dead volumes, which contribute to extra-column band broadening. The aim in packing a column is to minimize band broadening and peak distortion. Thus, it is desirable to pack a uniform bed with no cracks or

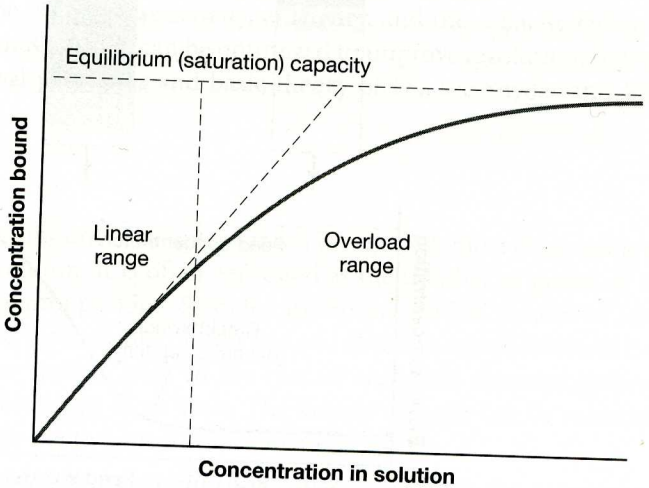


FIGURE 1.15. Representative sample adsorption isotherm depicting the equilibrium or saturation capacity and linear and "overload" regions. (Adapted, with permission, from PerSeptives Biosystems, Inc. 1996.)

TABLE 1.8. Effect of protein load on peak bandwidth for columns of varying internal diameters

Protein Load	Peak Volume (µl) for Column Dimensions			
(µg)	1.0 x 10 mm	2.1 x 3 mm	2.1 x 10 mm	4.6 x 10 mm
0.5	25	–	–	–
1.0	25	100	100	–
2.5	–	–	–	–
5.0	30	100	100	450
10.0	35	120	120	450
20	50	180	140	–
30	–	–	–	–
50	70	390	190	–
100	–	470	240	600
500	–	–	–	1200

Reproduced, with permission, from Simpson and Nice (1989 [©Wiley]).
Support: Brownlee RP-300 (column length = 10 cm).
Protein standard: bovine α-lactalbumin.
Gradient elution (0.15% TFA to 60% CH_3CN/0.12% TFA over 60 minutes) was used at equivalent linear flow velocities: 50 µl/min, 200 µl/min, and 1.0 ml/min for the 1.0-mm, 2.1-mm, and 4.6-mm I.D. columns, respectively.

channels and to maintain a uniform particle distribution within the column, by avoiding sizing or sorting of the particles during the packing procedure. Typically, rigid solids and hard gels are packed as densely as possible under high pressure, taking care to avoid fracturing the particles during the column-packing procedure. Soft gels, on the other hand, cannot be packed under high pressure, because they compress even at very low pressure.

Optimum packing procedures for the various soft gels are described by the manufacturers (see manufacturers' Web Sites). Common procedures for packing columns with rigid solids or hard gels include:

- "Dry-fill" packing (or the "tap-fill" dry-packing procedure) for rigid particles, with d_p >20 µm.
- "Wet-fill" column-packing methods (or slurry-packing procedures) for packing particles with d_p <20 µm.

High-pressure "wet fill" or slurry-packing methods for packing high-efficiency columns were very much in vogue 20 years ago, but today they are considered to be a technically difficult art form, requiring a high-pressure slurry-packing apparatus, and hence are not recommended outside the specialist laboratory. Moreover, the cost of prepacked columns has fallen significantly in recent years, making them affordable for most laboratories. For a discussion on high-pressure slurry-packing technologies, see Snyder and Kirkland (1979).

On the other hand, many simple column hardware systems are now commercially available for packing soft gels, and the procedures for packing such columns are relatively simple. Most modern columns are of the closed type—the basic component being a cylindrical tube, usually fabricated from stainless steel, glass, or plastic. The column is fitted with two end pieces, one fixed in place and the other an adjustable adapter that allows the column length and therefore the column volume to be varied (see Fig. 1.16). The end pieces are equipped with porous frits designed to retain the packing material while still allowing the mobile phase to pass through. The frits on fixed end pieces are often made of a plastic or metal screen, whereas on the adjustable adapters, the frits are fabricated from sintered metal, glass, or plastic, and designed to minimize abrasion of the chromatographic particles and to avoid clogging of the packed column. In most cases, frits are easily exchangeable and connect to a flow distributor designed to ensure an even flow across the entire column bed; the flow distributor, in turn, is connected to the column inlet and outlet.

On some columns, both end fittings are adjustable along the length of the column and can be fixed in place with a sliding seal mechanism. To ensure good sample separation (i.e., column effi-

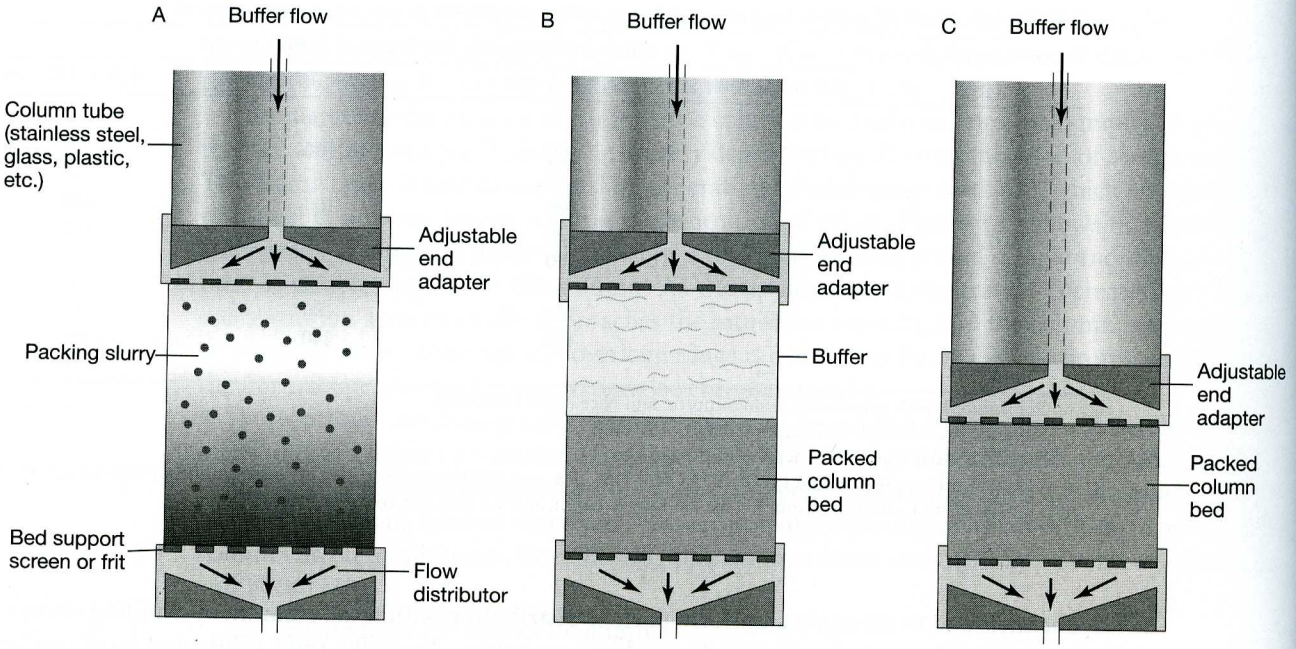


FIGURE 1.16. Basic elements of liquid chromatography column hardware and slurry packing. (*A*) Packing from the slurry settles to the bottom of the column. (*B*) Column bed is established beneath buffer. (*C*) Once column bed is consolidated, the adjustable end adapter is pushed down, so that it is flush with the top of the packed column. This eliminates any dead volume of buffer on top of the packed bed. (Adapted, with permission, from PerSeptives Biosystems, Inc. 1996.)

ciency), the column end fittings should be easily adjustable to prevent a void at the top of the bed, which is perhaps the most critical attribute contributing to poor sample resolution. If the column can be packed without any void spaces, then flow in the column can be either "up" or "down"; however, it is recommended that column flow be in the same direction in which the column was packed.

Column diameter terminology

The nature and dimensions of columns used in column chromatography can be divided into two broad categories (see Fig. 1.17).

- ***Packed columns.*** The stationary phase is affixed or bonded onto an inert solid matrix, and the particles are packed into a column. Packed columns can be further categorized based on their internal diameter (I.D.).

  Conventional columns have I.D.s >1.0 mm.

  Microbore columns have I.D.s in the range 0.5–1.0 mm.

  Capillary columns have very small I.D.s, <0.50 mm.

- ***Open tubular capillary columns.*** The stationary phase is bonded as a thin film directly onto the internal wall of a length of narrow-bore glass tubing. Two types of open tubular capillary columns are distinguished based on the manner in which the stationary phase contacts the column wall.

  Wall-coated open tubular capillary columns. The stationary phase is affixed directly to the column wall.

  Support-coated open tubular capillary columns. The walls are first coated with a layer of fine particulate material, which is then bonded with the stationary phase.

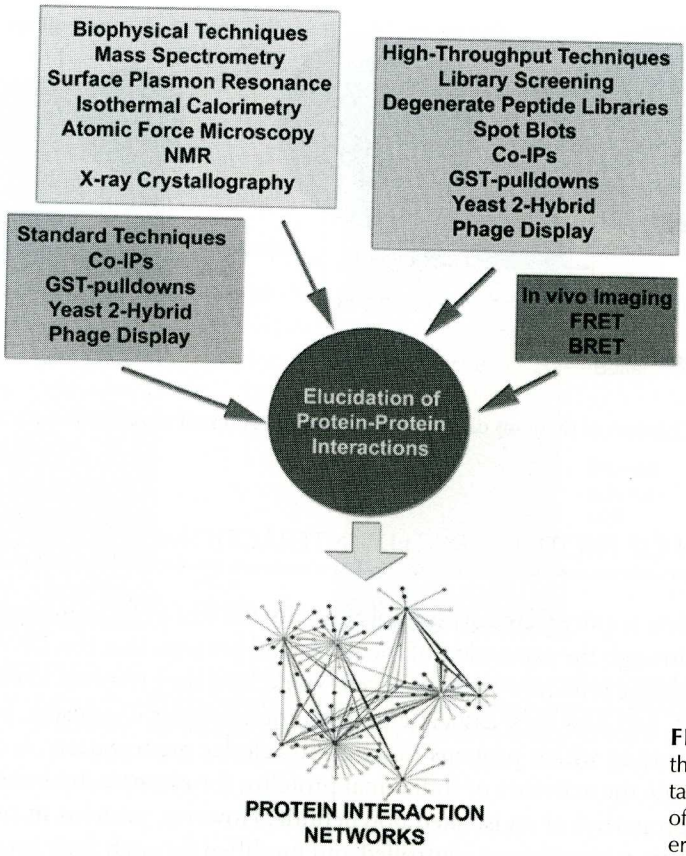


FIGURE 1.18. Flowchart illustrating how the juxtaposition biochemical and computational techniques leads to the elucidation of protein–protein interactions and the generation of protein interaction networks.

ferase (GST), which serve as probes to detect specific protein–protein interactions.

Once one or more protein–protein interactions have been identified, it is desirable to investigate the interaction(s) in detail, both in vitro and in vivo. Typically, the goals of this secondary analysis are (1) to obtain a mechanistic understanding of the proteins at the molecular level, (2) to understand the interaction's functional significance in vivo, and (3) to develop ways to specifically disrupt or perturb the interaction in vivo.

Until quite recently, the best way to determine the subcellular localization of a protein complex was by coimmunoprecipitation of the complex from a specific subcellular fraction of lysed cells (Chapter 12). In a direct extension of this approach, Chromatin ImmunoPrecipitation (ChIP) was developed as a means of localizing a DNA-binding protein or protein complex to a specific DNA sequence (Chapter 13). However, recent advances in fluorescence microscopy techniques, and the development of GFP and color variants as fusion–protein tags, have given rise to numerous other approaches to visualize protein–protein interactions in real time and in living cells. These include the use of fluorescence resonance energy transfer (FRET) and several variations of protein fragment complementation/bimolecular complementation (Chapter 15).

The genome-wide information obtained in the genomic and proteomic era has provided new opportunities that allow the identification of protein–protein interactions on a genome-wide level. In principle, it is now possible to screen every protein in the proteome for interactions with every other protein. However, the genomic/proteomic era has also provided new challenges. For example, the sometimes overwhelming body of data obtained can encourage a literal cataloging of interactions, without any reference to a specific biological question, context, mode of regulation, or any quantitation. Approaches that screen for protein–protein interactions in vivo may be best suited to addressing this problem in a single step. For example, using such methods, it should be

possible to compare networks of protein–protein interactions in cells grown under different conditions. Other approaches that screen for protein–protein interactions in vivo, such as variations on bimolecular complementation (Chapter 15), can be exploited in a similar way.

At present, meta-analyses of protein interaction data, together with other functional data, indicate that no single approach is optimal. Another problem of high-throughput, genome-wide approaches is that they tend to have relatively high rates of false positives, and likely higher rates of false negatives. To overcome this, methods are being developed to compare and integrate multiple high-throughput data sets (Lee et al. 2004), including functional data sets (Bouwmeester et al. 2004; Tewari et al. 2004). By comparing two or more data sets, each obtained by a distinct approach, one can be more confident of the physiological relevance of those interactions that appear in multiple data sets. One ultimate goal of these integrated high-throughput approaches is to build a dynamic network or map of interactions that properly predicts how a complex biological system maintains a steady state, or develops and differentiates toward a specific endpoint in response to a trigger. Several analytic tools are now available, including IM Browser, Cytoscape, and Osprey. Another goal of the genomic era is to be able to predict protein–protein interactions on the basis of primary protein sequence and by referring to structures of known protein–protein interactions. Concluding the final section of the book, Chapter 16 discusses how to access and exploit the vast amount of information available in comprehensive databases. A step-by-step example illustrates how this information, together with other resources including open access and online analytical programs, allows the investigator to develop and construct a broad network of both the physical and functional interactions for a given protein.

REFERENCES

Afeyan N.B., Gordon N.F., Mazsaroff I., Varady L., Fulton S.P., Yang Y.B., and Regnier F.E. 1990. Flow-through particles for the high-performance liquid chromatographic separation of biomolecules: Perfusion chromatography. *J. Chromatogr.* **519:** 1–29.

Amersham Biosciences. 1999. *Reversed-phase chromatography: Principles and methods handbook.* Amersham Biosciences, Buckinghamshire, United Kingdom.

Bonnerjea J.S., Oh S., Hoare M., and Dunnill P. 1986. Protein purification: The right step at the right time. *BioTechnology* **4:** 954–958.

Bouwmeester T., Bauch A., Ruffner H., Angrand P.O., Bergamini G., Croughton K., Cruciat C., Eberhard D., Gagneur J., Ghidelli S., et al. 2004. A physical and functional map of the human TNF-αNF-κB signal transduction pathway. *Nat. Cell Biol.* **6:** 97–105.

Burgess A.W., Metcalf D., Sparrow L.G., Simpson R.J., and Nice E.C. 1986. Granulocyte/macrophage colony-stimulating factor from mouse lung conditioned medium: Purification of multiple forms and radioiodination. *Biochem. J.* **235:** 805–814.

Cantor C.R. and Schimmel P.R. 1980. *Biophysical chemistry.* Part II. *Techniques for the study of biological structure and function,* Chapter 10. W.H. Freeman, San Francisco.

Catimel B., Ritter G., Welt S., Old L.J., Cohen L., Nerrie M.A., White S.J., Heath J.K., Demediuk B., Domagala T., Lee F.T., Scott A.M., Tu G.-F., Ji H., Moritz R.L., Simpson R.J., Burgess A.W., and Nice E.C. 1996. Purification and characterization of a novel restricted antigen expressed by normal and transformed human colonic epithelium. *J. Biol. Chem.* **271:** 25664–25670.

Cutler R.L., Melcalf D., Nicola N.A., and Johnson G.R. 1985. Purification of a multipotential colony-stimulating factor from pokeweed mitogen-stimulated mouse spleen cell conditioned medium. *J. Biol. Chem.***260:** 6579–6587.

Edman P. 1949. A method for the determination of the amino acid sequences in peptides. *Arch. Biochem.* **22:** 475–476.

Ford C.F., Suominen I., and Glatz C.E. 1991. Fusion tails for the recovery and purification of recombinant proteins. *Protein Expr. Purif.* **2:** 95–107.

Fountoulakis M. and Juranville J.F. 2003. Enrichment of low-abundance brain proteins by preparative electrophoresis. *Anal. Biochem.* **313:** 267–282.

Frolik C.A., Dart L.L, Meyers C.A., Smith D.M., and Sporn M.B. 1983. Purification and initial characterization of a type β transforming growth factor from human placenta. *Proc. Natl. Acad. Sci.* **80:** 3676–3680.

Görg A., Boguth G., Obermaier C., and Weiss W. 1998. Two-dimensional electrophoresis of proteins in an immobilized pH 4-12 gradient. *Electrophoresis* **19:** 1516–1519.

Görg A., Obermaier C., Boguth G., Harder A., Scheibe B., Wildgruber R., and Weiss W. 2000. The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **21:** 1037–1053.

Gospodarowicz D., Cheng J., Lui G.-M., Baird A., and Bohlent P. 1984. Isolation of brain fibroblast growth factor by heparin-Sepharose affinity chromatography: Identity with pituitary fibroblast growth factor. *Proc. Natl. Acad. Sci.* **81:** 6963–6967.

Graziano M.P., Moxham C.P., and Malbon C.C. 1985. Purified rat hepatic β_2-adrenergic receptor. *J. Biol. Chem.* **260:** 7665–7674.

Gygi S.P. and Aebersold R. 2000. Mass spectrometry and proteomics. *Curr. Opin. Chem. Biol.* **4:** 489–494.

Gygi S.P., Corthals G.L., Zhang Y., Rochon Y., and Aebersold R. 2000. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl. Acad. Sci.* **97:** 9390–9395.

Haga K. and Haga T. 1985. Purification of the muscarinic acetylcholine receptor from porcine brain. *J. Biol. Chem.* **260:** 7927–7935.

Han K.K. and Martinage A. 1992. Posttranslational chemical modification of proteins. *Int. J. Biochem.* **24:** 19–28.

Han K.K. and Martinage A. 1993. Post-translational modification of proteins. III. Current developments in analytical procedures of

identification and quantitation of post-translational chemically modified amino acid(s) and derivatives. *Int. J. Biochem.* **25:** 957–970.

Harlow E. and Lane D. 1999. *Using antibodies: A laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Hearn M.T.W. 1991. High-performance liquid chromatography of peptides and proteins: General principles and basic theory. In *High-performance liquid chromatography of peptides and proteins: Separation, analysis, and conformation* (ed. C.T. Mant and R.S. Hodges), pp. 95–122. CRC Press, Boca Raton, Florida.

Heil A., Müller G., Noda L., Pinder T., Schirmer H., Schirmer I., and von Zabern I. 1974. The amino acid sequence of sarcine adenylate kinase from skeletal muscle. *Eur. J. Biochem.* **43:** 131–144.

Heldin C.-H., Westermark B., and Wasteson A. 1981. Platelet-derived growth factor. *Biochem. J.* **193:** 907–913.

Henzel W.J., Tropea J., and Dupont D. 1999. Protein identification using 20-minute Edman cycles and sequence mixture analysis. *Anal. Biochem.* **267:** 148–160.

Hinds M.G. and Norton R.S. 1997. NMR spectroscopy of peptides and proteins. *Mol. Biotechnol.* **7:** 315–331.

Huber L.A., ed. 2003. Paper Symposium: Proteomics of cell organelles. *Electrophoresis* **21:** 3329–3527.

Jung E., Heller M., Sanchez J.C., and Hochstrasser D.F. 2000. Proteomics meets cell biology: The establishment of subcellular proteomes. *Electrophoresis* **21:** 3369–3377.

Klose J. 1975. Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik.* **26:** 231–243.

Koppel J., Boutterin M.C., Doye V., Peyro-Saint-Paul H., and Sobel A. 1990. Developmental tissue expression and phylogenetic conservation of stathmin, a phosphoprotein associated with cell regulation. *J. Biol. Chem.* **265:** 3703–3707.

Krishna R.G. and Wold F. 1993. Post-translational modification of proteins. *Adv. Enzymol. Relat. Areas Mol. Biol.* **67:** 265–298.

Krishna R.G. and Wold F. 1997. Identification of common post-translational modifications. In *Protein structure: A practical approach,* 2nd edition (ed. T.E. Creighton), pp. 91–116. IRL Press, Oxford University Press, United Kingdom.

LaVallie E.R. and McCoy J.M. 1995. Gene fusion expression systems in *Escherichia coli. Curr. Opin. Biotechnol.* **6:** 501–506.

Lee I., Date S.V., Adai A.T., and Marcotte E.M. 2004. A probabilistic functional network of yeast genes. *Science* **306:** 1555–1558.

Lottspeich F., Houthaeve T., and Kellner R. 1994. The Edman degradation. In *Microcharacterisation of proteins* (ed. R. Kellner et al.), pp. 117–130. VCH, Weinheim,. Germany.

Lowe C.R., Burton S.J., Burton N.P., Alderton W.K., Pitts J.M., and Thomas J.A. 1992. Designer dyes: 'Biomimetic' ligands for the purification of pharmaceutical proteins by affinity chromatography. *Trends Biotechnol.* **10:** 442–448.

Luche S., Santoni V., and Rabilloud T. 2003. Evaluation of nonionic and zwitterionic detergents as membrane protein solubilizers in two-dimensional electrophoresis. *Proteomics* **3:** 249–253.

Makrides S.C. 1996. Strategies for achieving high-level expression of genes in *Escherichia coli. Microbiol. Rev.* **60:** 512–538.

Marshak D.R. Kadonaga J.T., Burgess R.R., Knuth M.W., Brennan W.A. Jr., and Lin S.-H., eds. 1996. *Strategies for protein purification and characterization. A laboratory course manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

McPherson A. 1990. Current approaches to macromolecular crystallization. *Eur. J. Biochem.* **189:** 1–23.

McPherson A . 1997. Recent advances in the microgravity crystallization of biological macromolecules. *Trends Biotechnol.* **15:** 197–200.

Mikes O. and Coupek J. 1990. Organic supports. In *HPLC of biological macromolecules: Methods and applications* (ed. K.M. Gooding and F.E. Regnier), pp. 25–46. Marcel Dekker, New York.

Moritz R.L. and Simpson R.J. 1995. High-speed chromatographi separation of proteins and peptides for high sensitivity microse quence analysis. In *Methods in protein structure analysis* (ec M.Z. Atassi and E. Appella), pp. 27–38. Plenum Press, New Yorl

Nice E.C. and Catimel B. 1999. Instrumental biosensors: New per spectives for the analysis of biomolecular interactions. *BioEssay* **21:** 339–352.

Nice E.C. and Catimel B. 2000. High-performance liquid chromato graphic separations and equipment in peptide and protei analysis, miniaturization of. In *Encyclopedia of analytical chem istry: Applications, theory, and instrumentation* (ed. R.A Meyers), pp. 5823–5845. Wiley, New York.

Nicola A., Metcalf D., Matsumoto M., and Johnson G.R. 1983 Purification of a factor inducing differentiation in murine myelomonocytic leukemia cells—Identification as granulocyte colony stimulating factor. *J. Biol. Chem.* **258:** 9017–9023.

Nimmo G.A. and Cohen P. 1978. The regulation of glycogen metabolism. Purification and characterisation of protein phosphatase inhibitor-1 from rabbit skeletal muscle. *Eur. J. Biochem.* **87:** 341–351.

O'Farrell P.H. 1975. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250:** 4007–4021.

Parekh R.B. and Rohlff C. 1997. Post-translational modification of proteins and the discovery of new medicine. *Curr. Opin. Biotechnol.* **8:** 718–723.

PerSeptives Biosystems, Inc. 1996. *The busy researcher's guide to biomolecule chromatography.* Applied Biosystems, Foster City, California.

Qadri F. 1985. The reactive triazine dyes: Their usefulness and limitations in protein purifications. *Trends Biotechnol.* **3:** 7–11.

Rabilloud T. 2002. Two-dimensional gel electrophoresis in proteomics: Old, old fashioned, but it still climbs up the mountains. *Proteomics* **2:** 3–10.

Rabilloud T. 2003. Membrane proteins ride shotgun. *Nat. Biotechnol.* **21:** 508–510.

Rubinstein M., Rubinstein S., Familetti P.C., Miller R.S., Waldman A.A., and Pestka S. 1979. Human leukocyte interferon: Production, purification to homogeneity, and initial characterization. *Proc. Natl. Acad. Sci.* **76:** 640–644.

Sambrook J. and Russell D.W. 2001. *Molecular cloning: A laboratory manual,* 3rd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Schaller H.C. and Bodenmuller H. 1981. Isolation and amino acid sequence of a morphogenetic peptide from hydra. *Proc. Natl. Acad. Sci.* **78:** 7000–7004.

Scopes R.K. 1987. *Protein purification: Principles and practice,* 2nd edition. Springer-Verlag, New York.

Shu Z.Y. and Bi R.C. 1997. Optimizing protein crystallization. *Prog. Biochem. Biophys.* **24:** 396–401.

Simpson R.J. 2003. *Proteins and proteomics: A laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Simpson R.J. and Nice E.C. 1989. Strategies for the purification of sub-nanomole amounts of protein and poly-peptides for microsequence analysis. In *The use of HPLC in receptor biochemistry,* pp. 201–244. A.R Liss, New York.

Snyder L.R and Kirkland J.J. 1979. *Introduction to modern liquid chromatography,* 2nd edition. Wiley, New York.

Stanley P. 1989. Chinese hamster ovary cell mutants with multiple glycosylation defects for production of glycoproteins with minimal carbohydrate heterogeneity. *Mol. Cell. Biol.* **9:** 377–383.

Sulkowski E. 1985. Purification of proteins by IMAC. *Trends Biotechnol.* **3:** 1–7.

Tastet C., Lescuyer P., Diemer H., Luche S., Van Dorsselaer A., and Rabilloud T. 2003. A versatile electrophoresis system for the analysis of high- and low-molecular-weight proteins.

Electrophoresis **24:** 1787–1794.

Tatemoto K. 1982. Isolation and characterization of peptide YY (PYY), a candidate gut hormone that inhibits pancreatic exocrine secretion. *Proc. Natl. Acad. Sci.* **79:** 2514–2518.

Tewari M., Hu P. J., Ahn J. S., Ayivi-Guedehoussou N., Vidalain P.O., Li S.,Milstein S., Armstrong C.M., Boxem M., Butler M.D., et al. 2004. Systematic interactome mapping and genetic perturbation analysis of a *C. elegans* TGF-β signaling network. *Mol. Cell* **13:** 469–482.

Turner A.J. 1981. Scope and applications of dye-ligand chromatography. *Trends Biochem. Sci.* **6:** 171–193.

Ungar K.K. 1990. Silica as a support. In *HPLC of biological macromolecules: Methods and applications* (ed. K.M. Gooding and F.E. Regnier), pp. 3–24. Marcel Dekker, New York.

van Driel I.R., Stearn P.A., Grego B., Simpson R.J., and Goding J.W. 1984. The receptor for transferrin on murine myeloma cells one step purification based on its physiology and partial amino acid sequence. *J. Immunol.* **133:** 3220–3224.

Walker P.A., Leong L.E.-C., Ng P.W.P., Tan S.H., Waller S., Murphy D., and Porter A.G. 1994. Efficient and rapid affinity purification of proteins using recombinant fusion proteases. *Bio/Technology* **12:** 601–605.

Wang A.M. and Creasey A.A. 1985. Molecular cloning of the complementary DNA for human tumor necrosis factor. *Science* **228:** 149–154.

Wildgruber R., Harder A., Obermaier C., Boguth G., Weiss W., Fey S.J., Larsen P.M., and Gorg A. 2000. Towards higher resolution: Two-dimensional electrophoresis of *Saccharomyces cerevisiae* proteins using overlapping narrow immobilized pH gradients. *Electrophoresis* **21:** 2610–2616.

Yan S.C. and Grinnell B.W. 1989. Post-translational modification of proteins: Some proteins left to solve. *Trends Biochem Sci.* **14:** 264–268.

Zhang W., Zhou G., Zhao Y., White M.A., and Zhao Y. 2003. Affinity enrichment of plasma membrane for proteomics analysis. *Electrophoresis* **24:** 2855–2863.

Zugaro L.M., Reid G.E., Ji H., Eddes J.S., Murphy A.C., Burgess A.W., and Simpson R.J. 1998. Characterization of rat brain stathmin isoforms by two-dimensional gel electrophoresis-matrix assisted laser desorption/ionization and electrospray ionization-ion trap mass spectrometry. *Electrophoresis* **19:** 867–876.