TECHNIQUES IN MOLECULAR BIOLOGY

Volume 2

Edited by John M Walker and Wim Gaastra

TECHNIQUES IN MOLECULAR BIOLOGY

Volume 2

Edited by John M. Walker and Wim Gaastra

CROOM HELM
London & Sydney

Croom Helm, Provident House, Burrell Row,
Beckenham, Kent BR3 1AT
Croom Helm Australia, 44-50 Waterloo Road,
North Ryde, 2113, New South Wales

British Library Cataloguing in Publication Data

Techniques in molecular biology.
Vol. 2
1. Molecular biology—Technique
I. Walker, John M. II. Gaastra, Wim
574.8'8'028 QH506

ISBN 0-7099-3653-2
ISBN 0-7099-3673-7 Pbk

Typeset by Leaper & Gard Ltd, Bristol
Printed and bound in Great Britain by Mackays of Chatham Ltd, Kent

PREFACE

The last few years have seen the rapid development of new methodology in the field of molecular biology. New techniques have been regularly introduced and the sensitivity of older techniques greatly improved upon. Developments in the field of genetic engineering in particular have contributed a wide range of new techniques. In Volume 1, published in 1983, we introduced the reader to a selection of the more advanced analytical and preparative techniques which we considered to be frequently used by research workers in the field of molecular biology. In choosing techniques for Volume 1 we obviously had to be selective and were unable to cover as broad a spectrum of techniques as we would have liked. However, the production of Volume 2 has allowed us to develop the theme initiated in Volume 1 and also expand to include a wider range of subject areas. As with Volume 1, the majority of chapters relate to nucleic acid methodology, but we have also covered immunological methodology and protein purification techniques that were not included in Volume 1. Obviously, we see Volume 2 as simply a continuation of Volume 1. As with Volume 1, a knowledge of certain basic biochemical techniques and terminology has been assumed. However, since many areas of molecular biology are developing at a formidable rate and constantly generating new terminology, a glossary of terms has been included.

Each chapter aims to describe both the theory and relevant practical details for a given technique, and to identify both the potential and limitations of the technique. In some cases more detailed practical descriptions are given to indicate the scale at which the work is carried out. Each chapter is written by authors who regularly use the technique in their own laboratories. This book will prove useful to final year undergraduate (especially project) students, postgraduate research students and research scientists and technicians who wish to understand and use new techniques, but who do not have the necessary background for setting up such techniques. Although lack of space precludes the description of in-depth practical detail (such as buffer compositions, etc.), such information is available in the references cited.

J.M. Walker
W. Gaastra

CONTENTS

1 MICROMETHODS IN PROTEIN CHEMISTRY

Elaine L.V. Harris

1. Introduction

The technical advances made in the past decade in both protein and nucleic acid chemistry have enabled considerable development in the role of today's protein chemist. The techniques introduced in the late 1950s and early 1960s for studying protein structure were designed for dealing with abundant proteins readily available in micromolar (that is, 10^{-6} molar), or greater quantities (for example, haemoglobin, histones and the digestive enzymes). In recent years the interest of the molecular biologist has turned to the less abundant proteins, which can only be purified in nanomolar, or picomolar quantities (that is, 10^{-9} or 10^{-12} molar), such as some biologically active peptides (growth factors, hormones or neurotransmitters), and their receptors. The fundamental principles of those early techniques, such as chromatography, gel electrophoresis, protein sequencing and amino acid analysis, still apply to today's techniques. However, the speed and sensitivity have been increased many-fold due to changes in both instrumentation and methods of detection. Improvements in the matrices used for chromatography have led to the advent of HPLC (high performance liquid chromatography) and FPLC (fast protein liquid chromatography). These two techniques have proved extremely valuable for rapid, high resolution separation of nanomolar, or smaller, quantities of proteins, peptides and amino acids. Probably the most outstanding advance has been made in protein sequencing, where the advent of the gas-phase sequencer has led to a 10000–100000 fold increase in sensitivity over Edman's original method, although the chemistry remains essentially unchanged.

In the late 1970s some scientists believed that the rapid advances being made in nucleic acid chemistry heralded the decline and fall of protein chemistry. However, this has proved not to be the case, and today protein

chemistry still plays a key role in many areas of biochemical research. In fact biochemical research nowadays requires a multidisciplinarian approach, drawing together protein chemistry, recombinant DNA technology, DNA synthesis, peptide synthesis and immunology. The rapid, simple methods now available for DNA sequencing enable the complete amino acid sequence of a protein to be obtained more efficiently, albeit indirectly, by these methods rather than by direct protein sequencing. However, if a particular gene encoding one of the less abundant proteins is to be studied, the protein must first be purified, and either antibodies made to it and/or partial amino acid sequence determined. The antibodies can then be used to detect recombinant DNA clones expressing the protein. Alternatively, the partial amino acid sequence is screened for five to seven contiguous amino acids which when translated into the genetic code have the least ambiguity. Mixtures of 14 to 20 base oligonucleotide probes can be rapidly synthesised, covering all the possible sequences predicted from the amino acid sequence. These probes are subsequently used to detect recombinant DNA clones containing the gene. In many instances, once the DNA sequence has been obtained knowledge of the amino acid sequence is still required to determine the boundaries and the reading frame of the gene, and also to confirm that the correct gene has been sequenced. Many proteins are modified post-translationally by proteolytic processing, disulphide bond formation, glycosylation or phosphorylation, and protein sequencing is the only method of identifying the position of these modifications. The current methods of protein sequencing are quantitative, allowing accurate determination of both subunit stoichiometry in complex multisubunit enzymes, and purity of protein or peptide preparations. Thus the purity of synthetically made peptides is established beyond doubt by sequencing, using either the gas-phase sequencer or fast atom bombardment (FAB) mass spectrometry. Protein sequencing is also essential to the biotechnology industry, where particularly for therapeutic use absolute purity of a protein produced by recombinant DNA technology must be established prior to use.

In the following sections only the more recent advances in protein chemistry techniques will be discussed.

Perhaps the motto of a protein chemist using microtechniques should be 'Cleanliness is next to godliness!' When dealing with proteins and peptides in low nanomolar, or picomolar quantities significant contamination can be introduced by dirty glassware, dust or from the hands of the worker. Thus, all glassware should be thoroughly cleaned with chromic acid, or a good laboratory detergent, rinsed thoroughly with high quality distilled water, and stored away from dust. Towards the end of a purification in particular it is advisable to wear plastic, talc-free gloves when handling vessels containing the protein. Contaminants can also be introduced by poor quality reagents, thus all reagents should be of analytical, or better quality. Poor quality reagents may also cause blocking of the *N*-terminus of the protein,

rendering it resistant to protein sequencing using the Edman chemistry. Another problem frequently encountered with low protein concentrations is non-specific binding to surfaces of glassware, etc. The amounts lost vary from protein to protein depending on the net charge and/or hydrophobicity of the protein, and can be of the order of a few micrograms (μg). These losses can be minimised by siliconising glassware, or by using containers made of polypropylene.

2. Quantitative Protein Assays

Purification of a particular protein is best monitored by the change in the ratio of the concentration of the protein to total protein concentration (that is, the specific activity of the protein). For enzymes and biologically active polypeptides their concentrations can usually be easily quantified using highly sensitive assays based on their functions. Many assays are available for determining the concentration of total protein.[1] Clearly these assays must be non-destructive, or highly sensitive, so that significant proportions of the purified protein are not destroyed in the assay. Some of the more useful assays for microprotein chemistry are discussed below.

Non-destructive Assays

Absorbance at either 280 nm or 200–230 nm can be used as a non-destructive protein assay. Absorbance at 280 nm is dependent on the amount of tryptophan and tyrosine present, and therefore shows considerable variation between proteins. The absorbance at 200–230 nm is due to the peptide bond (as well as tyrosine, tryptophan, phenylalanine, histidine, methionine and cysteine), and is therefore subject to less variation between proteins than is the absorbance at 280 nm. Protein concentrations of 50–1000 μg/ml are required for measurement at 280 nm, while at 200–230 nm the absorbance is higher, and concentrations of 5–200 μg/ml can be detected. Many compounds, such as nucleic acids or detergents, interfere with absorption measurements particularly at the lower wavelengths.

Destructive Assays

The Lowry assay, a colorimetric assay for proteins based on the blue colour developed by interaction between both tyrosine and tryptophan residues and a mixture of phosphomolybdic and tungstic acids, was much used in the past. However, this assay has been largely superseded by the Coomassie brilliant blue assay initially developed by Bradford,[2] and subsequently modified by Read and Northcote.[3] The Coomassie blue binding assay is simple and rapid and is both more sensitive and less prone to interference by other compounds (such as sucrose or Tris) than is the Lowry assay. Protein solutions containing 1–140 μg (10–1400 μg/ml) are mixed with the reagent solution and the absorbance at 595 nm measured after 5 min.

Glass or plastic cuvettes must be used since the dye binds to quartz cuvettes. Either bovine serum albumin or gammaglobulin is normally used as the standard. The mechanism of Coomassie blue binding is not yet fully understood; it has been suggested that it binds to either lysine or arginine. The response in the assay varies from protein to protein, and thus the value obtained for the amount of protein present in the assay is not absolute, but is relative to the protein used as a standard. Commercial preparations of the reagent solution are available (such as BioRad Laboratories, USA; Pierce Chemicals, USA), which show less batch-to-batch variation than do solutions made in the laboratory. An adaptation of this assay for use in microtitre plates has recently been reported,[4] with a sensitivity of less than 0.5 μg for bovine serum albumin.

For more sensitive detection, assays using o-phthalaldehyde (opa) or fluorescamine are used.[1] Opa is preferable since it is less expensive, less pH-dependent and stable in aqueous solutions. The reaction which occurs between opa, thiol reagent and amino acids is shown in Figure 1.1. A thiol reagent such as 2-mercaptoethanol must be present for the reaction to proceed. The protein solution, containing 100 ng or more, is mixed with the opa reagent in the presence of 2-mercaptoethanol, and after 15 min the fluorescence is measured, using an excitation wavelength of 340 nm and an emission wavelength of between 440 and 455 nm. Since opa binds only to the amino terminus and lysine side chains, the response in this assay again varies from protein to protein. However, this variation can be overcome by acid hydrolysis of the protein to its constituent amino acids prior to addition of the opa/thiol reagent. Prior acid hydrolysis also has the advantage of increasing the sensitivity a further ten-fold, due to the increase in amino groups. Unfortunately, primary amines such as Tris, and high concentrations of ammonia interfere with this assay.

Amino Acid Analysis

Amino acid analysis has several uses in microprotein chemistry:

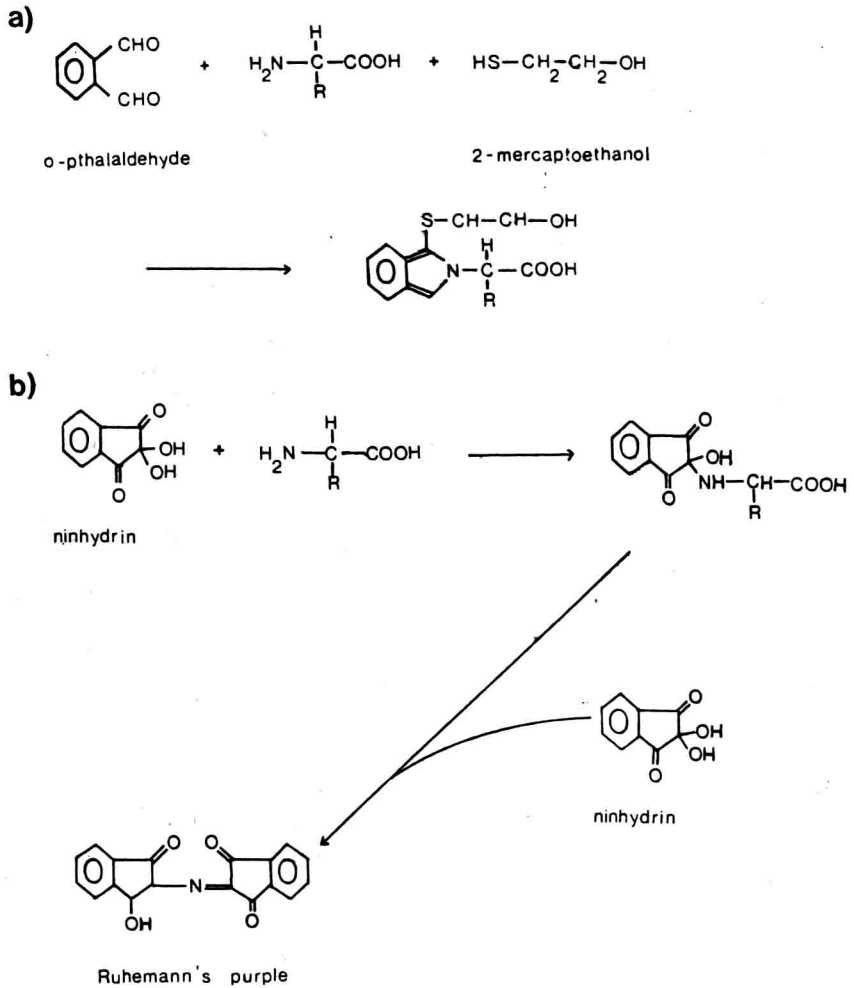
(1) determination of absolute protein concentration;
(2) determination of composition; and
(3) determination of purity of proteins and peptides.

The latter two uses are discussed later in this chapter.

For amino acid analysis the protein or peptide must first be hydrolysed to its constituent amino acids. This is usually achieved by incubating the sample under vacuum in a sealed tube with 6M hydrochloric acid at 110°C for 4–24 h. Longer hydrolysis times are required to break peptide bonds between pairs of some hydrophobic amino acids, such as valine, leucine and isoleucine. However, longer hydrolysis times partially destroy threonine and serine.

The method originally developed by Moore and Stein[5] used a sul-

Figure 1.1: (a) Reaction of o-pthalaldehyde and 2-mercaptoethanol with an amino acid; (b) Reaction of ninhydrin with an amino acid producing Ruhemann's purple which absorbs at 570 nm



phonated cation-exchange resin to separate the amino acids, which were then detected with ninhydrin. Many modifications of this method have been made since, resulting in increased sensitivity, faster analysis and increased resolution. Most commercially available amino acid analysers use ion-exchange resins to resolve the amino acids and either ninhydrin or opa/thiol to detect them. The reactions which occur between amino acids and ninhydrin or opa/thiol are shown in Figure 1.1. Detection limits for postcolumn derivatisation with ninhydrin are between 1 and 10 nmol while with opa the limits can be as low as 50 pmol.

A more sensitive technique, which is gaining popularity, is precolumn derivatisation with opa, in the presence of thiol, followed by separation of the amino acid derivatives by reverse-phase HPLC.[6] Fluorescent detection limits are between 1 and 10 pmol, and typical run times are less than 60 min. The thiol–opa–amino acids are unstable, particularly those formed with lysine and glycine. The derivatisation must therefore be carried out under controlled conditions of time and temperature, immediately prior to injection onto the HPLC column. This is best achieved using an autosampler capable of either mixing the opa–thiol reagent with the sample solution in a tube immediately prior to injection, or 'sandwiching' the sample between aliquots of opa–thiol reagent in the injection loop. A typical trace obtained for a standard mix of amino acids is shown in Figure 1.2.

During hydrolysis tryptophan is destroyed, cysteine, methionine and tyrosine are partially oxidised, and asparagine and glutamine are converted to aspartic and glutamic acids. For determination of protein concentration these effects do not significantly alter the results. More serious errors can be introduced when using opa detection methods, since proline and hydroxyproline are not detected by this reagent. Thus, if the absolute protein concentration is to be obtained, primary amines must be introduced into these imino acids. This is usually achieved by oxidation with chloramine T, or hypochlorite to open the imino ring.[7,8] Primary amines (such as Tris), detergents, and high concentrations of salts in the sample can interfere with amino acid analysis by precolumn or postcolumn derivatisation. It should also be stressed that, particularly with the more sensitive opa detection methods, cleanliness is of the utmost importance, since significant amounts of amino acid contaminants will be picked up from dust and from the workers' hands! Tubes for hydrolysis should preferably be baked at very high temperatures (450°C), to destroy any contaminating amino acids or proteins prior to use, and stored dust-free.

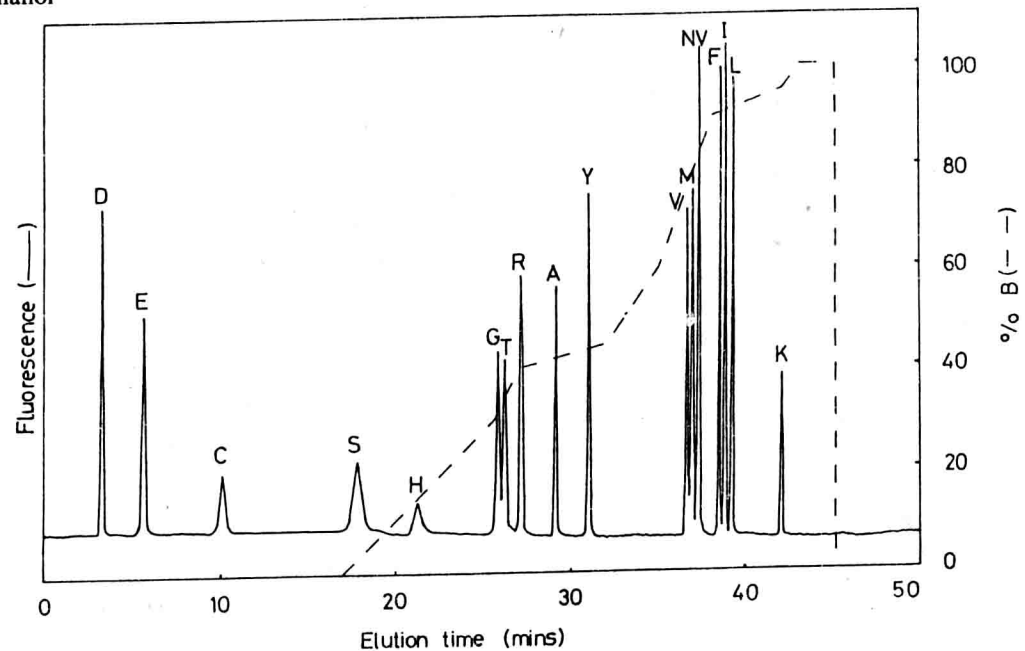
The weight of protein which was hydrolysed can be calculated by addition of the products obtained for each amino acid for the following equation:

number of moles determined × molecular weight

3. Protein Purification

The ultimate goal for a protein purification is 100% purity of the protein of interest, with the highest yield. Unfortunately, particularly when dealing with low protein concentrations, losses occur at every stage of the purification due to degradation by proteases, denaturation, and absorption to glassware and chromatography gels. Thus, although conventional ion-exchange and gel filtration chromatography are still used, particularly in the early stages of purification, they have been largely superseded by tech-

Figure 1.2: Amino acid analysis of a standard mix of amino acids using precolumn derivatisation with o-pthalaldehyde and 2-mercaptoethanol



Note: After addition of the opa-thiol reagent to the amino acid solution the mixture was injected onto a C_{18} reverse-phase HPLC column (Rainin Microsorb, 15 cm × 4.5 mm internal diameter). The flow rate was 1.5 ml/min and the buffers used were: (a) 54.3 mM propionic acid/76 mM disodium hydrogen phosphate (adjusted to pH 6.5 with sodium hydroxide): acetonitrile (1840 : 160, v/v); (b) water : methanol : acetonitrile : dimethyl sulphoxide (40 : 600 : 500 : 60, v/v/v/v)). The eluate was monitored for fluorescence with an excitation wavelength of 335 nm and an emission wavelength of 455 nm. The amount of each amino acid injected was 10 pmol. The standard one-letter code is used to label the peaks (D = aspartic acid; E = glutamic acid; C = cysteic acid; S = serine; H = histidine; G = glycine; T = threonine; R = arginine; A = alanine; Y = tyrosine; V = valine; M = methionine; NV = norvaline [included as an internal standard]; F = phenylalamine; I = isoleucine; L = leucine; K =

niques with greater resolution (such as affinity chromatography, HPLC and FPLC) which enable purification to be achieved in fewer steps. As mentioned in the previous section, each stage is monitored by specific activity; purity is said to be attained when no further increase in specific activity is observed in subsequent purification procedures. Additional methods of monitoring purity are HPLC and sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) (see Volume 1). A protein which shows a single peak on HPLC and a single silver-stained band after SDS-PAGE is, in most cases, pure.

Affinity Chromatography

Affinity chromatography (see Chapter 3) exploits either the biological function of a protein, or its individual chemical or immunological structure, enabling extremely selective purification of the protein to be achieved. The protein is specifically and reversibly absorbed to a compound known to bind to the protein, which is irreversibly attached to a solid support (matrix). These binding compounds fall into three categories as outlined below:

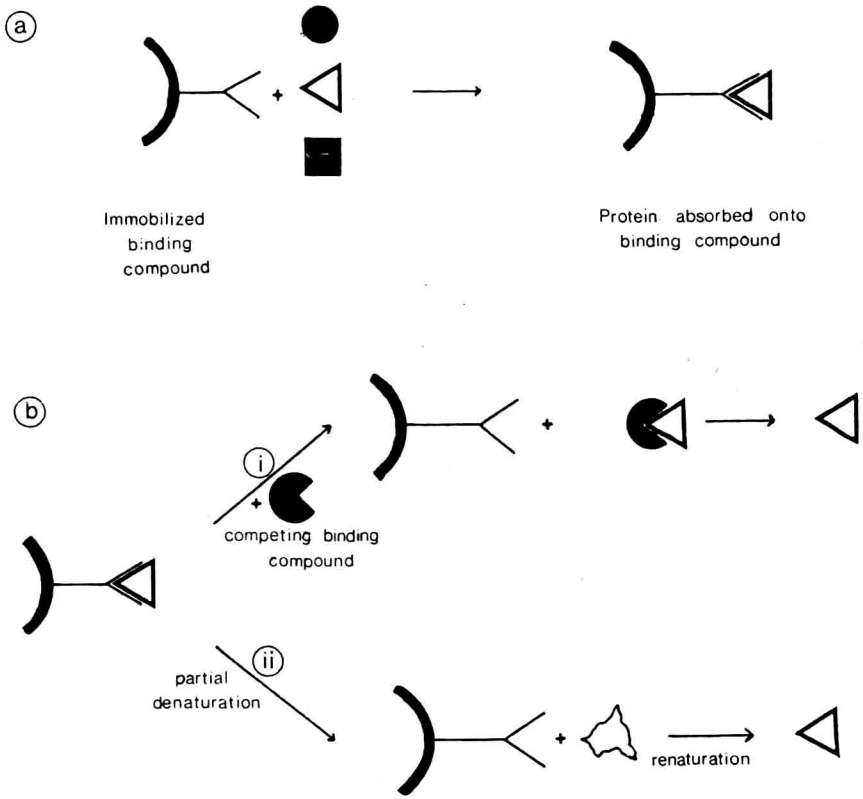
(1) *Cofactors.* Those in common use are NAD, NADP, ATP and CoA. Matrices with immobilised triazine-based dyes are also used to purify cofactor-dependent enzymes, as well as a variety of other proteins. The most commonly used dye is Cibacron Blue F3G-A and a commercial matrix based on this dye is available (Blue Sepharose, Pharmacia, Sweden).
(2) *Ligands.* Ligands with high affinity for the protein of interest can be immobilised on solid supports to make highly specific affinity matrices. Either the native ligand can be used or an analogue of the native ligand.
(3) *Antibodies.* If antibodies, preferably monoclonal, to the protein of interest are available these can be immobilised on solid supports to make highly selective affinity matrices.

After the protein is bound the matrix is washed to remove unbound contaminants, and then the protein is eluted by addition of a competing binding compound, or by changing the buffer conditions to promote dissociation from the binding compound. Figure 1.3 shows a schematic representation of a purification using affinity chromatography. The purification achieved is typically of the order of several thousand-fold and the recovery of activity is frequently high. Affinity chromatography is covered in more detail in Chapter 3.

HPLC and FPLC

Improvements in the nature of column packing materials have led to the advent of HPLC (high performance liquid chromatography) and FPLC (fast protein, polypeptide, polynucleotide liquid chromatography). These

Figure 1.3: Schematic representation of affinity purification

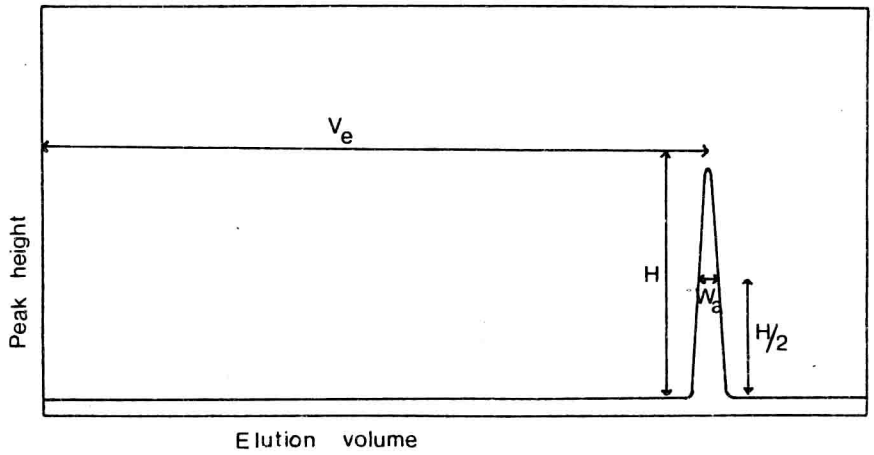


Note: A compound to which the protein of interest is known to bind is immobilised on a solid support (such as Sepharose beads), usually by covalent attachment. (a) A solution containing the protein is mixed with the immobilised compound and the protein becomes absorbed onto the solid support through specific interaction with the binding compound. (b) After washing to remove unbound contaminants the protein is eluted either: (i) specifically by addition of a binding compound which competes for the binding site of the protein; (ii) by alteration of the buffer conditions (such as pH, ionic strength or addition of denaturants) thereby causing the protein to dissociate from the immobilised compound

improvements have yielded smaller rigid beads, with greater uniformity in size and shape (typically 3 or 5 μm ± 25% for HPLC columns and 10 μm ± 2% for the mono beads used in some FPLC columns), than those of the conventional soft gels (for example, Sephadex beads are 80 μm ± 50%, and Trisacryl beads are 60 μm ± 30%). The small, uniform beads can be packed in a column with minimum spaces between beads, thus minimising broadening of eluted molecules caused by diffusion of the molecules within these spaces. Therefore, for a given size column a considerably higher resolution in separation is obtained with the smaller, uniform beads. The reso-

lution, or efficiency of a column is measured in theoretical plates (N) and is calculated as shown in Figure 1.4. Typical efficiencies for HPLC and FPLC column packings are 30000 theoretical plates per metre, whereas the conventional soft gels give only 2000–5000 theoretical plates per metre. HPLC and FPLC columns can therefore be smaller than columns packed with the conventional soft gels, and still achieve higher resolution. Since the columns are smaller, separations can be completed in considerably shorter times by HPLC or FPLC than by soft gels. As a consequence of the tighter packing achieved with the HPLC and FPLC packings higher pressures must be used, than with the larger soft gels, to achieve a given flow rate. Thus, both the column packings and the HPLC and FPLC pumps have been designed to be compatible with these higher pressures. HPLC columns are typically run at 1500–3000 psi (10–20 MPa), whereas FPLC columns with their larger, more porous beads are run at between 150–1500 psi (1–10 MPa) (conventional soft gels are used at 3 psi (0.02 MPa)). Both HPLC and FPLC columns are usually packed by their manufacturers. HPLC columns are usually stainless steel, whereas the resins used for FPLC, because they are run at lower pressures, are packed in glass

Figure 1.4: Calculation of theoretical plates for a chromatography column



$$\text{Theoretical plates (N)} = 5.54 \times \left(\frac{V_e}{W_a}\right)^2$$

Note: The peak used to obtain the various parameters for the calculation should be one which elutes near the end of the chromatography run (for example, the peak obtained from acetone injected onto a gel filtration column). V_e is the elution volume (in either ml or min) of the peak from the start of the chromatography run, and W_a is the width of the peak (in the same units as V_e) at half its height. Theoretical plates per metre can be calculated by dividing N by the length of the column (in metres)

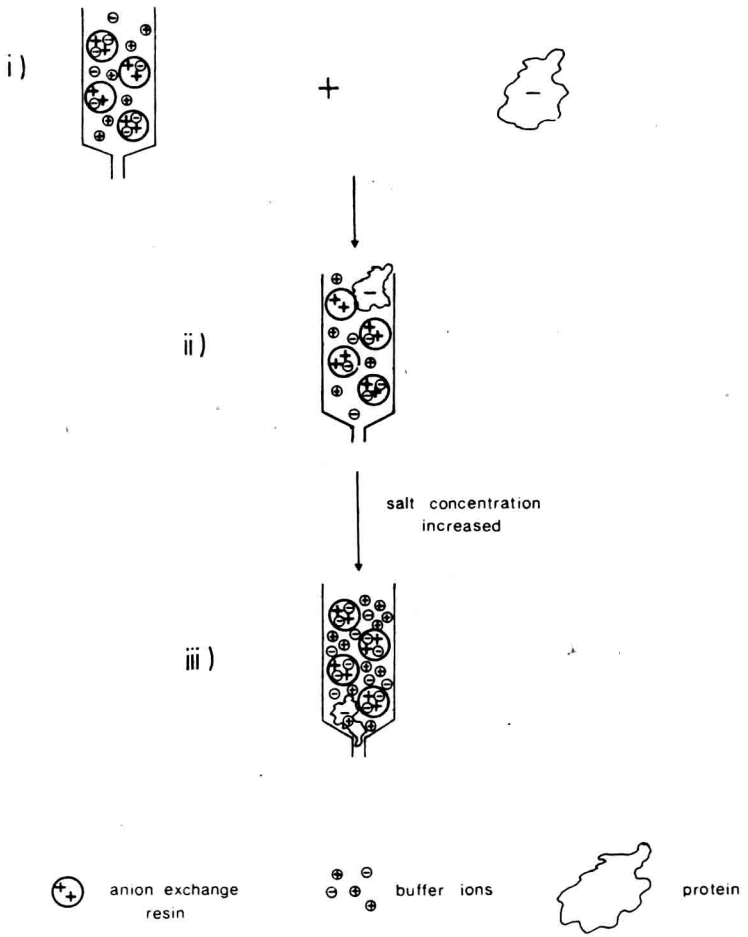
columns. Glass columns have the advantage of allowing easy visual inspection of the resin at all times. The wettable parts of the FPLC system are totally inert (PTFE, glass and titanium) allowing prolonged use of high salt concentrations, which can cause corrosion of the stainless steel used in HPLC systems. The use of inert materials also ensures that metal ions, which might inactivate some enzymes, do not leach from the system.

Both HPLC and FPLC have gained popularity due to their speed, versatility, high resolution and sensitivity. The maximum loading capacity of most HPLC and FPLC columns is between 1 and 100 mg total protein. Larger capacity columns are available, but are very costly. Therefore, crude extracts are usually partially purified by ion-exchange, gel filtration or affinity chromatography on conventional soft gels, prior to HPLC or FPLC. However, the speed of HPLC and FPLC runs, typically less than 60 min, allows several runs to be performed in less time than one run using conventional gels (with which typical runs usually take longer than one day). This speed also enables rapid screening of different columns and different buffer conditions to achieve optimum purification. The growing range of matrices available for HPLC and FPLC enables proteins and peptides to be separated on the basis of charge, size or hydrophobicity, as described below.

Ion Exchange (IEC) and Chromatofocusing Chromatography Both these techniques separate proteins on the basis of their charge, and are complementary. Figures 1.5 and 1.6 show simplified representations of the chromatography on ion-exchange and chromatofocusing resins.

An ion-exchange resin consists of charged groups covalently attached to an insoluble matrix; ions of the opposite charge (counter-ions) present in the mobile phase bind to these charged groups. The counter-ions can be displaced by other counter-ions present in proteins or buffers. Either positive or negative groups (such as diethylaminoethyl or carboxymethyl groups, respectively) can be attached to the resin giving, respectively, anion or cation exchangers. Anion exchangers are used at pH values above the isoelectric point (pI) of the protein, where the net charge of the protein is negative, while cation exchangers are used at lower pH values. Bound proteins are eluted either by a gradient of increasing concentrations of counter-ions (such as NaCl) or by changing the pH so that the net charge of the protein is changed either to the opposite charge, or to neutrality. Many IEC columns, both anion and cation exchangers, are available for both HPLC and FPLC. These are either based on silica or hydrophilic resins, the latter being compatible with pH values between 2 and 10, whereas silica based resins can only be used between pH 2.5 and 7.5. Proteins with molecular weights up to one million can be separated on the resins currently available. Typical separations by HPLC or FPLC IEC take 30–40 min, but trial runs to determine optimum conditions can be completed in under 5 min. The maximum loading capacities are typically 25–

Figure 1.5: Principles of ion-exchange chromatography



Note: In the example shown here the resin is an anion exchanger, and has numerous covalently attached positively charged groups (i) Negative ions present in the starting buffer interact with these charged groups. The protein solution is also equilibrated in the starting buffer and is then applied to the column (ii) Negative charges on the protein compete for the positively charged groups on the resin and therefore bind to the resin; uncharged proteins or those with a net positive charge pass straight through the column (iii) The salt concentration of the buffer is increased, thereby increasing the number of ions present. These ions shield the charges on both the resin and the protein thus decreasing the interaction between them and causing the protein to elute from the column