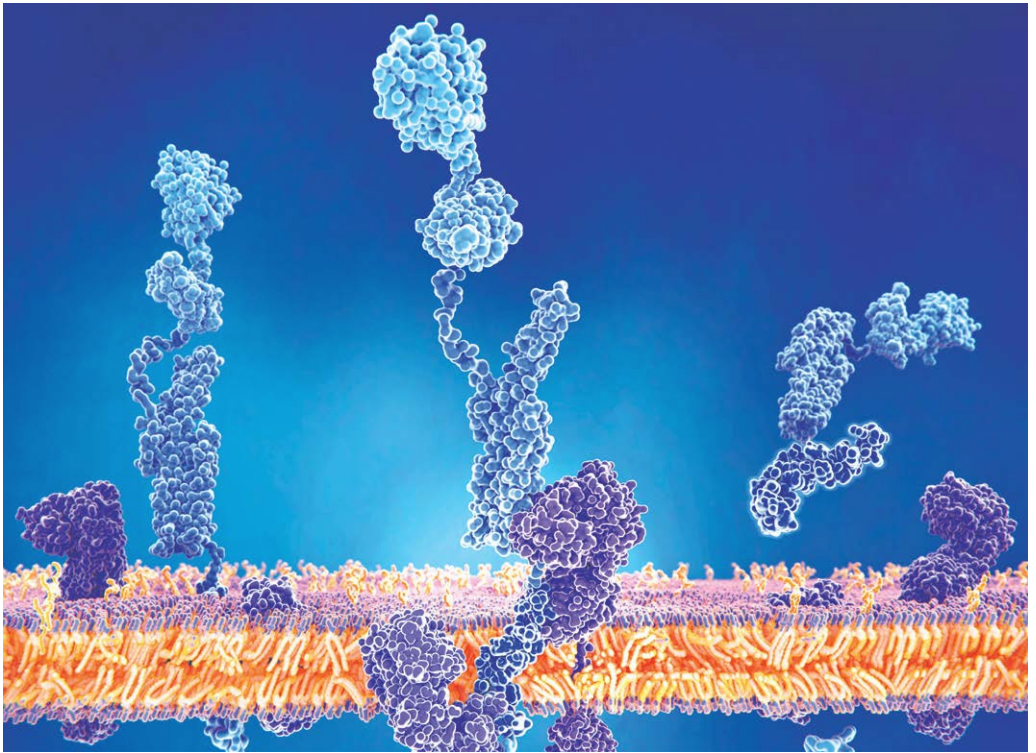


Chromatography Solutions

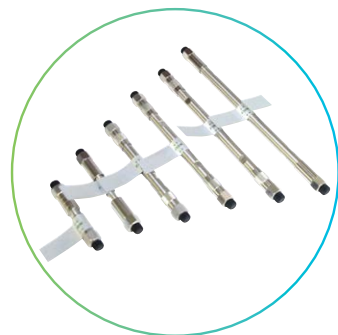
# Avantor® ACE® peptide and protein guide



Avantor® ACE®

# A guide to the analysis and purification of proteins and peptides by reversed-phase HPLC

by Dave Carr



## Table of contents

Introduction  
Page 3

Mechanism of protein/  
peptide retention  
Page 5

Column characteristics  
Page 7

Mobile phase  
Page 12

Detection  
Page 19

Effect of temperature  
Page 20

Reversed-phase HPLC-MS  
Page 22

Peptide mapping  
Page 26

Protein oxidation  
Page 32

Protein deamidation  
Page 34

Disulphide bond  
determination  
Page 37

Protein/Peptide  
purification  
Page 40

Proteomics  
Page 45

References  
Page 49

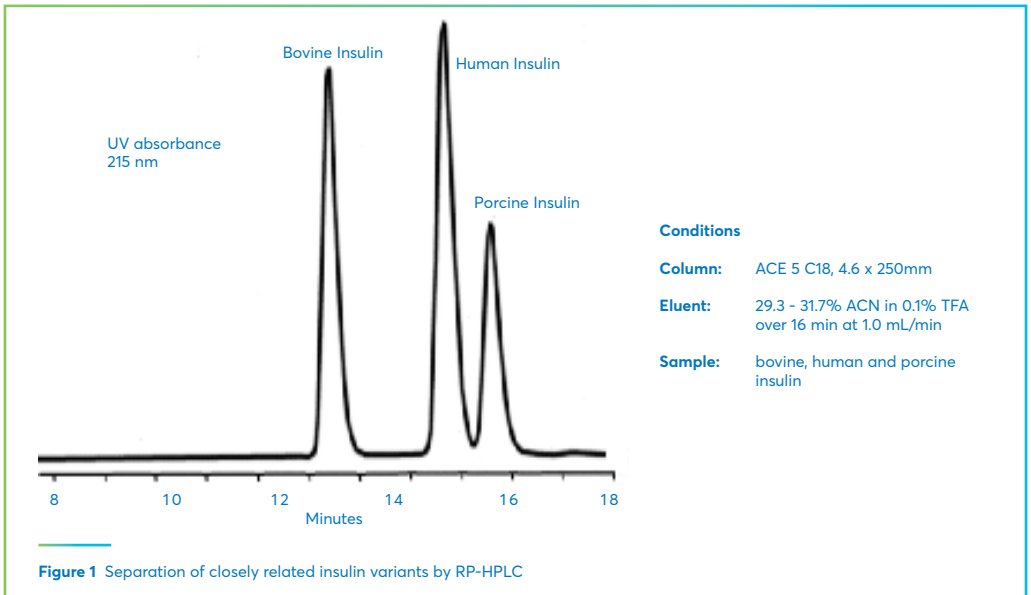
Appendix: Column  
Characteristics  
Page 53

Appendix: Amino Acid  
Characteristics  
Page 54

# Introduction

Reversed-phase HPLC has become an essential tool in the separation and analysis of proteins and peptides. It is widely used in the biotechnology industry to characterize protein therapeutic products and to analyze these for product identity and impurities. Reversed-phase HPLC plays a vital role in the separation of peptides from digested proteomes prior to protein identification by mass spectrometry. It is also used to purify many proteins and peptides during investigative studies and is used for large scale purification of protein therapeutic drugs.

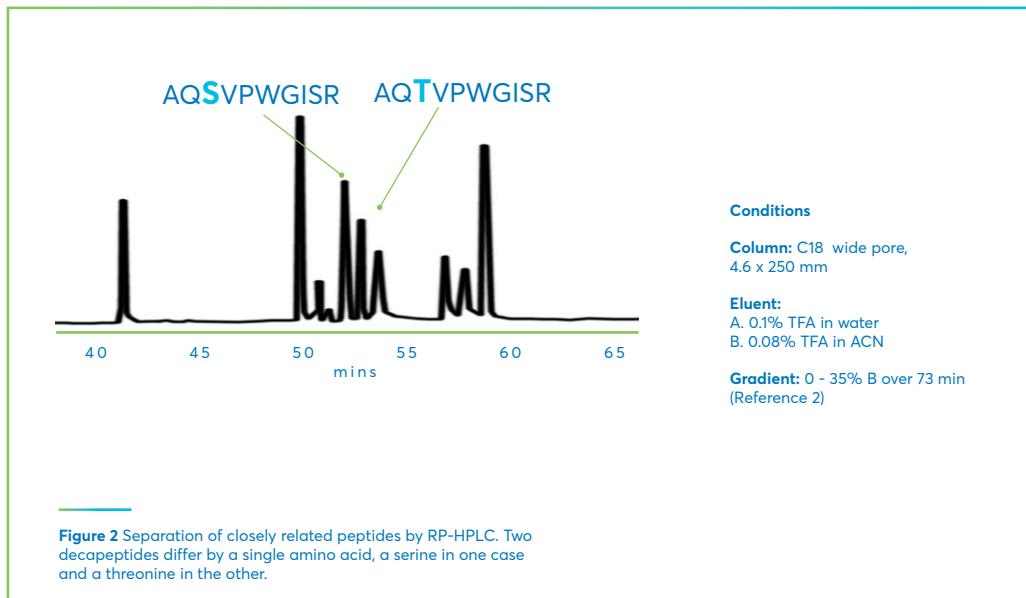
Reversed-phase HPLC has found a central role in protein studies because of its versatility, sensitive detection and its ability to work together with techniques such as mass spectrometry. Most of all, however, reversed-phase HPLC is widely used because of its ability to separate proteins of nearly identical structure. As illustrated by the separation of bovine, human and porcine insulin variants (Figure 1), reversed-phase HPLC is able to separate very similar proteins. Bovine and human insulin differ by only three amino acids and are well separated. Bovine insulin has an alanine at



residue 8 and a valine at residue 10 on the insulin 'a' chain and an alanine at residue 30 of the 'b' chain. Human insulin has a threonine at residue 8 and an isoleucine at residue 10 on the 'a' chain and a threonine at residue 30 of the 'b' chain. Porcine and human insulin differ by just one amino acid (porcine has an alanine at residue 30 of the 'b' chain; human insulin has a threonine at that position) and are baseline resolved. In another example rabbit insulin was separated from human insulin even though they differ by one amino acid, threonine in place of serine (Reference 1).

The high resolving capability of reversed-phase HPLC extends to peptides as well. In Figure 2, two peptides are shown to be well resolved although differing by only a single amino acid, serine versus threonine.

It is this high resolution capability that is the foundation for the widespread use of reversed-phase HPLC for the separation of proteins and peptides.

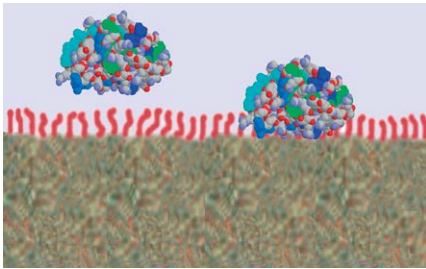


# Mechanism of protein/ peptide retention

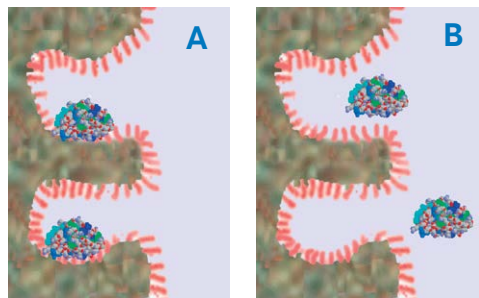
In reversed-phase HPLC the particle surface is very hydrophobic due to the chemical attachment of hydrocarbon groups to the surface (wavy red lines in Figure 3). Proteins are retained by the adsorption of a face of the protein (termed the "hydrophobic foot") to the hydrophobic surface (Figure 3). Since proteins are large compared to the thickness of the hydrophobic surface, only a portion of the protein adsorbs to the hydrophobic surface. Much of the protein lies above the surface and is in contact with the mobile phase. The net interaction caused by this hydrophobic adsorption is very strong resulting in the protein remaining adsorbed to the surface (Figure 4A) until a specific concentration of organic solvent is reached, at which time the protein desorbs from the surface

and elutes from the column (Figure 4B). Although there is some interaction with the surface as the protein moves down the column after the initial adsorption/desorption, further interactions are minor and do not contribute to separation. Separation is accomplished by the single adsorption/desorption process. The concentration of organic modifier required to desorb the protein is highly specific and is a function of the size of the hydrophobic foot. For further details see Reference 3.

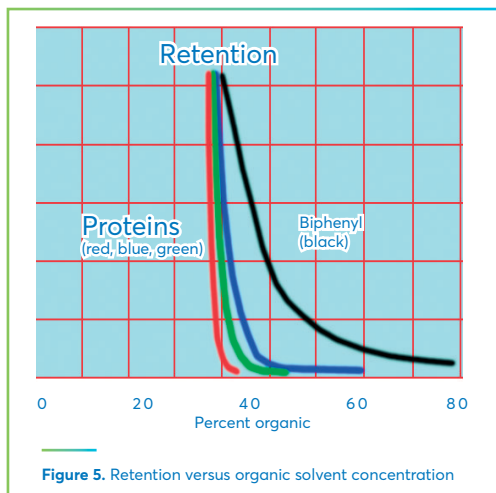
The adsorption/desorption retention mechanism leads to protein retention behaviour that is different than with small molecules in reversed-phase HPLC. While small molecules change



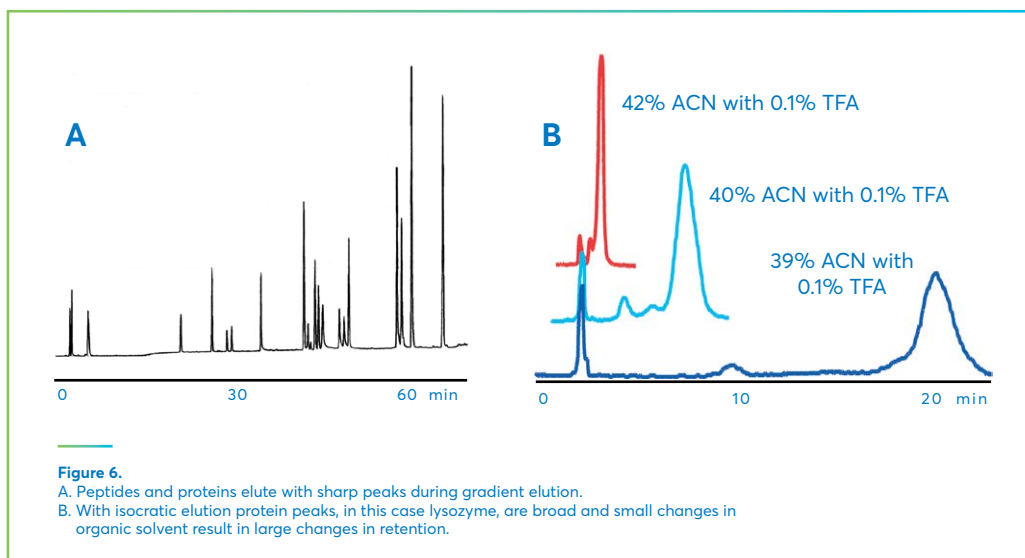
**Figure 3.** The relatively large protein molecule is adsorbed onto the hydrophobic surface by means of the hydrophobic effect



**Figure 4.** Proteins entering the column adsorb to the hydrophobic surface near the top of the column (A) and remain adsorbed until the organic modifier concentration reaches a specific concentration, when the protein desorbs from the surface (B).



retention slowly with changes in organic solvent concentration (Figure 5, biphenyl), the retention of proteins changes abruptly once the required concentration of organic solvent is reached, resulting in a rapid change in retention (Figure 5, proteins). This results in the sharp peaks usually seen with proteins and peptides (Figure 6A). The large change in retention with small changes in organic solvent concentration means that isocratic elution is seldom useful with proteins because peaks become broad and small changes in organic solvent concentration result in large changes in protein retention (Figure 6B).

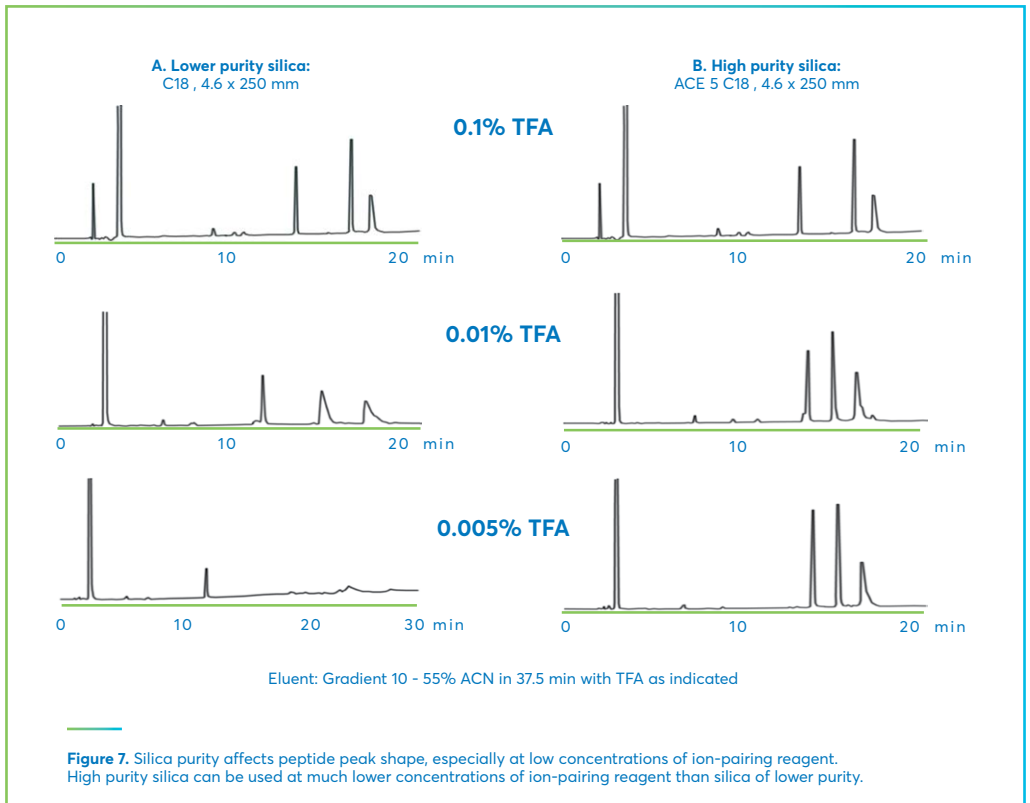


# Column characteristics



**Particles.** Proteins and peptides are separated by interacting with the hydrophobic surface of particles packed in columns. The particles in the column are usually made of silica because silica is physically robust, it is stable under most solvent conditions (except at pH greater than 6.5) and silica can be made into spherical particles of various sizes with pores of different diameters.

**Silica purity.** The purity of the silica used in HPLC columns is important in separation performance. Metal ion impurities cause peak tailing and loss of resolution as shown in Figure 7A (0.01 and 0.005% TFA). Silica with metal impurities (Figure 7A) requires the use of high concentrations of an ion-pair reagent, trifluoroacetic acid (TFA), to maintain good peak shape. The use of low concentrations

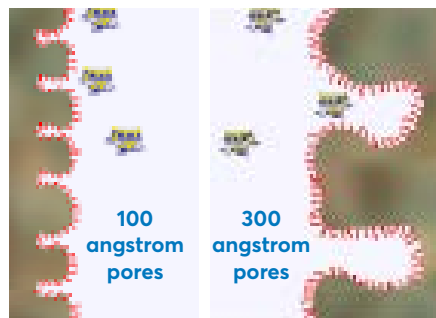


of TFA results in poor peak shape and loss of resolution. With high purity silica (Figure 7B), TFA concentrations as low as 0.005% result in good peptide peak shape. This is especially important in LC-MS because TFA causes a signal reduction when using the electrospray interface. Low concentrations of TFA result in better detection signal in LC-MS (see pages 22 - 25)

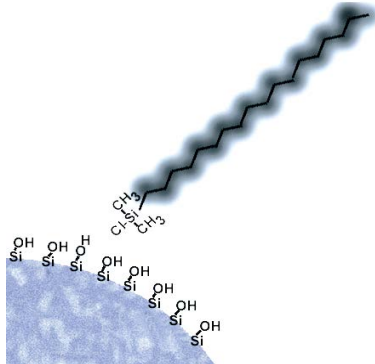
**Pore Diameter.** Using small pore (~ 100 angstrom) silica normally used in reversed-phase HPLC results in inferior protein separations. Wide pore silica (~ 300 angstrom diameter) gives much better separations of proteins (Reference 4). As illustrated in Figure 8, proteins cannot enter small pores, leaving only the very small exterior surface for separations to occur. Wide pore silica allows proteins and even larger peptides to enter the pore and fully interact with the surface, thus

resulting in better peak shape and resolution. Wide pore silicas are universally used today for protein separations. Small peptides, such as those resulting from protease digests, can enter the pores of small pore silica and interact with the surface and therefore small pore silicas may be used for separating protein digests. However wide pore silicas also separate peptides well and result in different selectivity and resolution.

**Hydrophobic surface.** The silica is modified with a hydrocarbon molecule in order to create a hydrophobic surface. A chlorosilane with an attached hydrocarbon chain, such as octadecylchlorosilane, is reacted with the silica (the surface of which consists of polar silanol groups) to attach the hydrocarbon to the silica surface (Figure 9). The organosilane molecules do not react with every silanol on the silica surface



**Figure 8.** Small pore (~ 100 angstrom) particles commonly used in reversed-phase HPLC (left) do not permit most proteins to enter the pores, thus limiting surface interaction. Particles with wide (~300 angstrom) pores (right) allow proteins to enter and interact with the hydrophobic surface.



**Figure 9.** The hydrophobic surface is formed by the chemical attachment of hydrophobic ligands to the silica surface by means of an organochlorosilane.



because of steric hindrance and a significant number of polar silanols remain on the surface. A process called “end-capping” whereby a small organosilane is subsequently reacted with the silica surface, further reduces the number of polar silanol groups.

**Choosing the separation surface.** The chemistry used to modify the silica surface allows various organic groups to be attached to the silica. The most common modification is attachment of a linear, aliphatic eighteen carbon chain resulting in a “C18” or ODS type column (Figure 10A). As shown, the organochlorosilane reacts with many of the silanols, but not all. This creates a fairly thick layer of hydrocarbon on the surface. It is to this thick hydrocarbon layer that proteins and peptides adsorb.

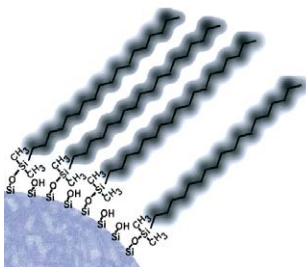
The C18 column is particularly useful for the separation of peptides less than 3,000 daltons and is usually the column of choice for the separation of peptides resulting from protease

digestion of proteins (see pages 26 - 31) as well as the separation of natural and synthetic peptides. A less hydrophobic phase results from the attachment of butyl (C4) groups to the silica surface (Figure 10B). The butyl phase is most suitable for protein separations but can also be used to separate large or hydrophobic peptides.

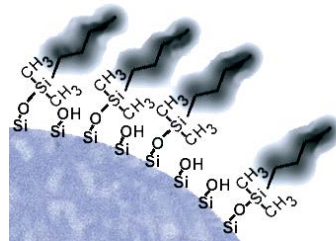
Proteins can be separated with a C18 column, however some proteins have poor peak shape or tailing peaks on C18 columns and C4 columns are recommended for protein separations.

Other columns used for polypeptide separations include phenyl phases (Reference 6), which are similar to C4 columns in hydrophobicity, and polar embedded or polar endcapped columns which enhance polar interactions between peptides and the particle surface. This may result in different selectivity for peptides.

**Peptide selectivity.** Column selectivity for peptides is affected by the nature and characteristics of



**Figure 10. A.** C18 hydrophobic phase is best suited to separation of peptides smaller than ~ 2000-3000 daltons



**B.** C4 hydrophobic phase is best suited to separation of peptides greater than ~ 3000 daltons and proteins

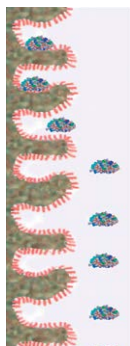
the bonded phase as well as the underlying silica surface. Different reversed phase columns may offer different peptide selectivity. In particular:

- The amount of phase ("carbon load") on the silica affects selectivity. When less hydrocarbon is attached to the surface (lower carbon load), polar silanols may affect the separation more than on columns with higher carbon load, thus resulting in different selectivity.
- Different manufacturing processes yield silica with different properties, which can affect peptide selectivity.

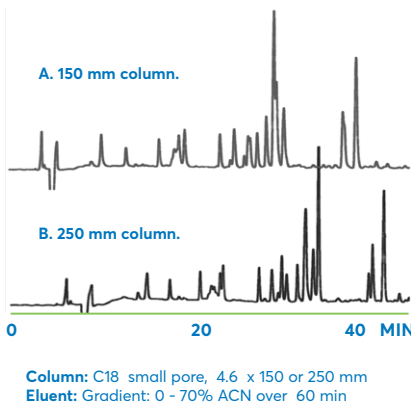
**Column length.** The more interaction small molecules have with the particle surface the better the resolution and long columns give higher resolution than short columns. Proteins, however, adsorb near the top of the

column, are subsequently desorbed and do not interact appreciably with the particle surface after being desorbed (Figure 11). Although data suggests that some interaction does occur between a protein and the particle surface, the interaction is not selective and does not contribute to resolution between proteins. **Column length is not important in protein separations and short columns separate proteins as well as long columns.**

Since peptides interact less strongly with the hydrophobic reversed-phase surface than proteins, length appears to play a greater role in peptide and protein digest separations. As illustrated in Figure 12, longer columns usually result in better resolution for peptides than shorter columns. **Fifteen or twenty-five centimetre columns are recommended for peptide separations.**



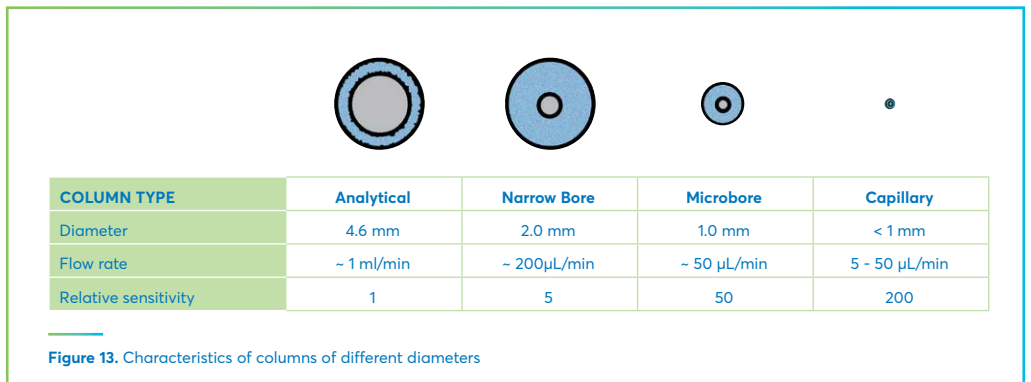
**Figure 11.** Proteins adsorb and desorb near the top of the column. Once desorbed, proteins interact minimally with the hydrophobic phase, thus longer columns do not increase resolution with proteins as they do with small molecules.



**Figure 12.** In contrast to proteins, peptides are usually better resolved on longer columns.

**Column Diameter.** The standard internal diameter of analytical HPLC columns is 4.6 mm. These columns are best run at flow rates of ~1 ml/min. Smaller bore columns are commercially available and are used for specific reasons and purposes. Narrow bore columns (~2 mm i.d.) are run at flow rates of ~ 200 microlitres/ min and thus use less solvent than 4.6 mm i.d. analytical columns. Narrow bore columns also exhibit about five times the sensitivity of standard analytical columns. This is because the amount of solvent passing through the detector per minute is lower, resulting in a higher concentration of protein or peptide in the peak. Concentration dependent detectors, such as the UV detector and electrospray mass spectrometer, show greater sensitivity with smaller bore columns.

Microbore columns operate at flow rates of ~50 microlitres/minute and therefore their use results in even better sensitivity, about 50 times that of an analytical column. Finally, capillary columns operating at 1 - 50 microlitres/minute flow rates exhibit even higher relative sensitivity, about 200 times the sensitivity of an analytical column. Microbore and capillary columns, however, require specialized or optimized instruments because of the flow rates being used and the greater significance of extra-column dead volume. With the introduction of UHPLC instrumentation (low extra volume), we have seen a move to narrower columns. Column characteristics are summarized in Figure 13 and also in the Appendix. See Reference 5 for details regarding the use of small bore columns.



**Figure 13.** Characteristics of columns of different diameters

# Mobile phase

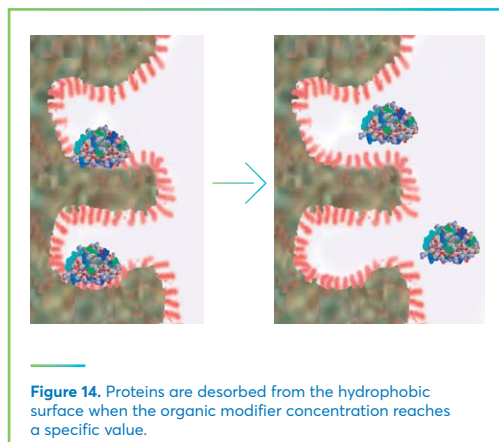
Proteins are desorbed from the hydrophobic surface with an organic solvent (Figure 14). When, during the solvent gradient, the amount of organic solvent reaches a precise concentration which is unique to each protein, the protein desorbs from the hydrophobic surface and continues down the column and elutes from the column.

**Acetonitrile.** The organic solvent most commonly used in polypeptide reversed-phase chromatography is acetonitrile. Acetonitrile is used because it:

- Is volatile and easily removed from the sample.
- Has low viscosity and thus low back pressure.
- Is quite transparent to low wavelength UV light.
- Has a long history of successful separations.

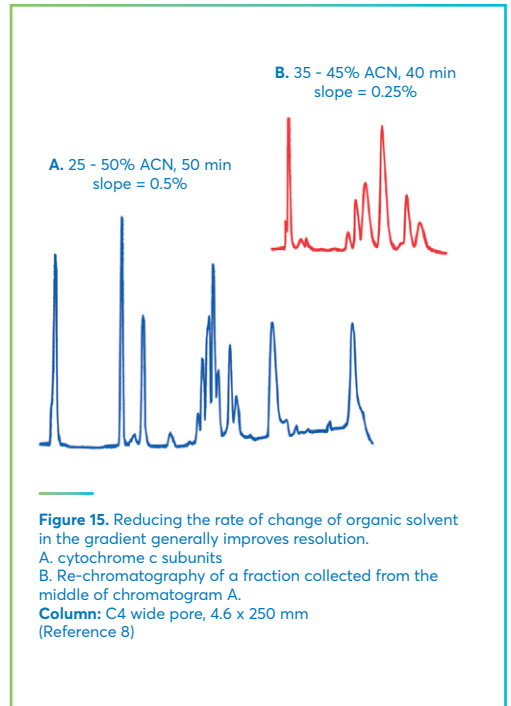
**Isopropanol.** Isopropanol plays a particular role in polypeptide chromatography. While seldom used as the sole organic modifier because of its high viscosity (and consequent high back pressure), it is useful to improve recovery of some polypeptides, particularly very hydrophobic proteins. In such cases isopropanol is added at a constant concentration of 1 - 5 percent to enhance recovery or elution of hydrophobic polypeptides.

**Other organic modifiers.** Organic modifiers such as methanol or ethanol are seldom used except for very hydrophobic proteins (Reference 7). Ethanol is also used for large scale process purification of proteins because of its low toxicity.



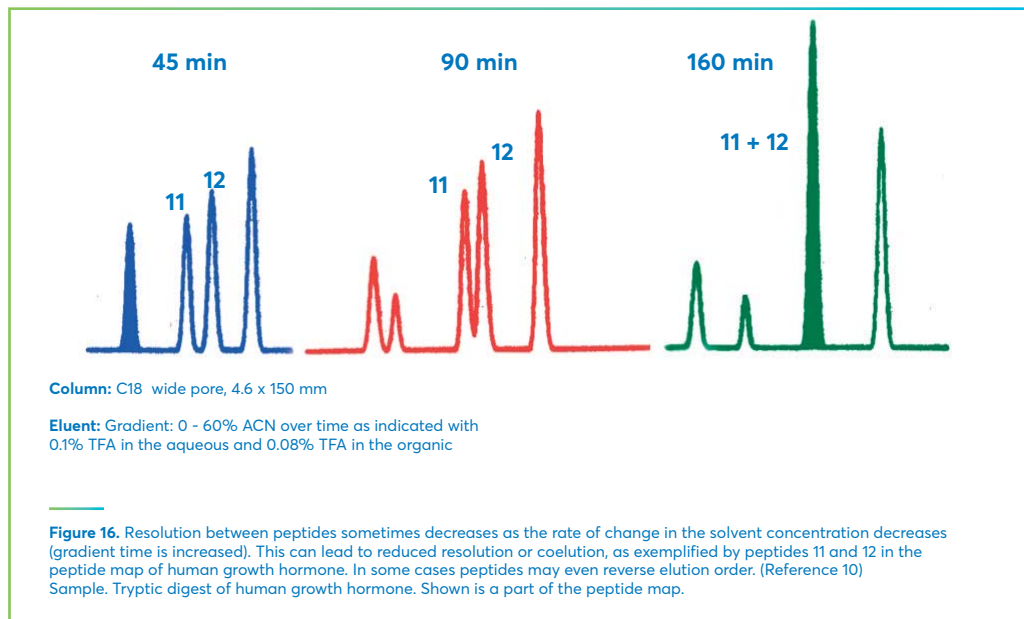
**Gradient elution.** Polypeptides are almost always eluted using a solvent gradient where the relative concentration of organic solvent is slowly increased during the separation. Proteins and peptides elute when the concentration of the organic modifier rises to the precise concentration required for desorption. As illustrated in Figure 15, the slower the rate of change of organic modifier concentration (the gradient slope), the better the resolution of these protein subunits. In this example a slope of 0.25% per minute resulted in significantly improved resolution compared to a slope of 0.5% per minute. Figure 1 (Page 3) shows the separation of insulin variants which were separated with a solvent change of 0.15% per minute. A gradient slope as low as 0.05% per minute may be used for maximum resolution.

**A reduction of the gradient slope to improve resolution must be balanced with the need for keeping analysis time as short as possible. However, adjusting the gradient slope is important in optimizing resolution of proteins and peptides.**



Peptides sometimes exhibit a peculiar behaviour as the gradient slope is decreased (Figure 16). Instead of improving resolution as expected, resolution is sometimes reduced leading to coelution and, sometimes, even a reversal of elution order. In the example in Figure 16, peptides 11 and 12 are best separated with a 45 minute gradient. When the gradient time is extended to 90 minutes, resolution between peptides 11 and 12 decreases and, with a gradient of 160 minutes, the

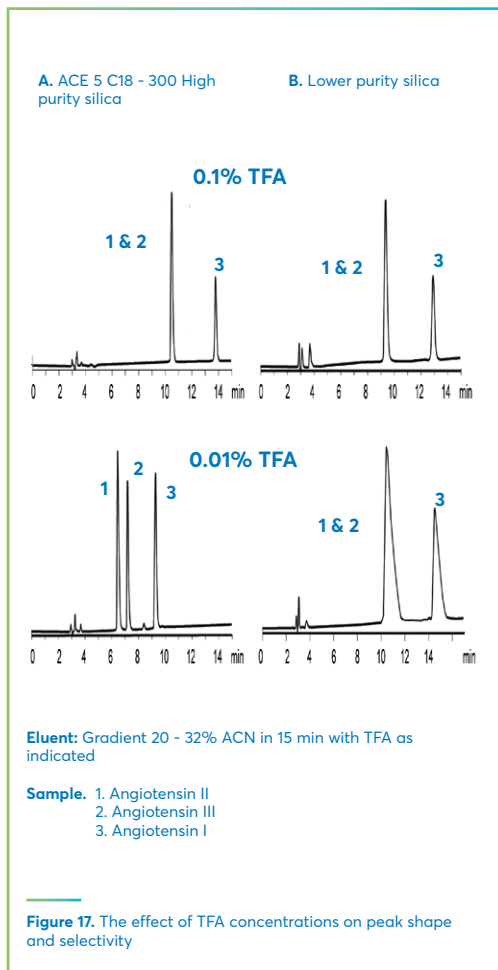
two peptides coelute. This retention behaviour is due to effects of the peptide-surface interaction, and a change in selectivity is observed. It is, therefore, important when developing peptide separations, particularly when separating protease digest samples, to observe changes in resolution as the gradient slope is decreased. If resolution decreases instead of increasing then the slope must be carefully optimized to maximize overall resolution.



The reversed-phase chromatography of proteins and peptides requires a reagent, called an “ion-pair”, added to the mobile phase in order to achieve good peak shape. It is believed that metal impurities on the silica surface are responsible for poor protein/peptide peak shape in the absence of an ion-pair reagent.

**Trifluoroacetic acid.** The most commonly used ion-pair reagent is trifluoroacetic acid (TFA). TFA added to the mobile phase at a concentration of ~0.1% results in good peak shape on most columns (Figure 17). Lowering the concentration of TFA improves LC-MS detection sensitivity (see pages 22 - 25) but may result in poor peak shape on silica columns because of impurities on the silica surface (Figure 17B). High purity silica columns, however, can be used at very low concentrations of TFA (Figure 17A - 0.01% TFA).

**The TFA concentration may affect peptide selectivity.** In Figure 17A (high purity silica), two peptides coelute with a TFA concentration of 0.1% but are resolved with a TFA concentration of 0.01%. (In Figure 17B - lower purity silica - the two peptides remain unresolved). Generally, varying the concentration of TFA from 0.2% to 0.01% may improve the selectivity of peptide separations on high purity silicas. The concentration of TFA may also be increased to solubilize hydrophobic proteins or peptides, however some columns may not be stable at TFA concentrations higher than 0.1%.

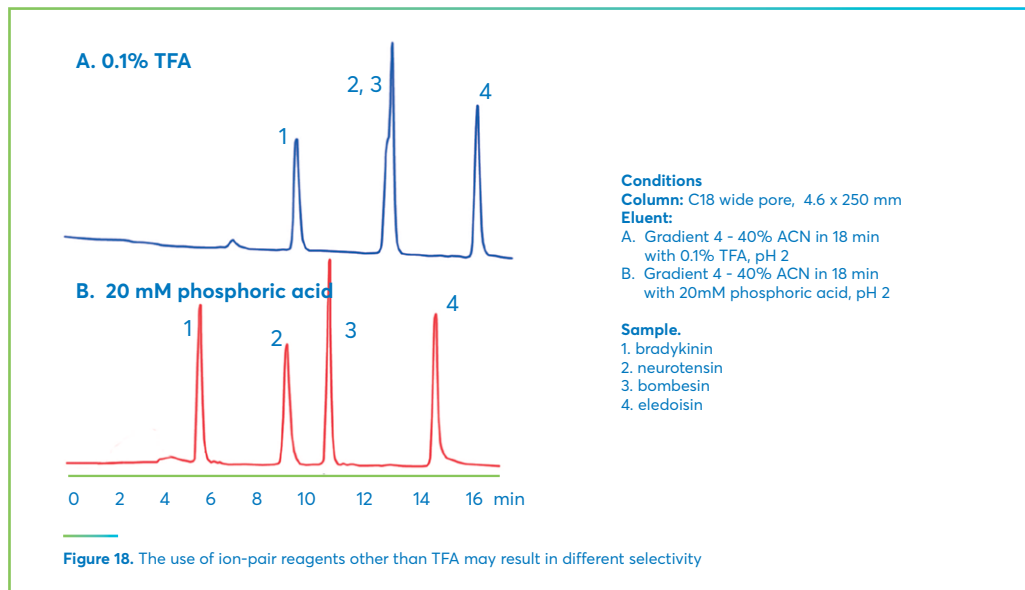


**Alternative ion-pair reagents.** Although TFA is by far the most commonly used ion-pair reagent, other reagents such as phosphoric acid and heptafluorobutyric acid (HFBA) are sometimes used in protein/peptide separations.

As illustrated in Figure 18, phosphoric acid can sometimes separate peptides that are not separated by TFA. Phosphate is used at concentrations around 20-30 mM, pH 2 - 2.5. Phosphate buffers have also been shown to offer better separations of some proteins than

TFA. While phosphate buffer is usually used at low pH like TFA, phosphate buffers can also be adjusted to higher pH, providing an opportunity to change selectivity and resolution (see page 15). The principal drawback of using phosphate as an ion-pair reagent is that it is not volatile, making it difficult to remove from the peptide and non LC-MS compatible.

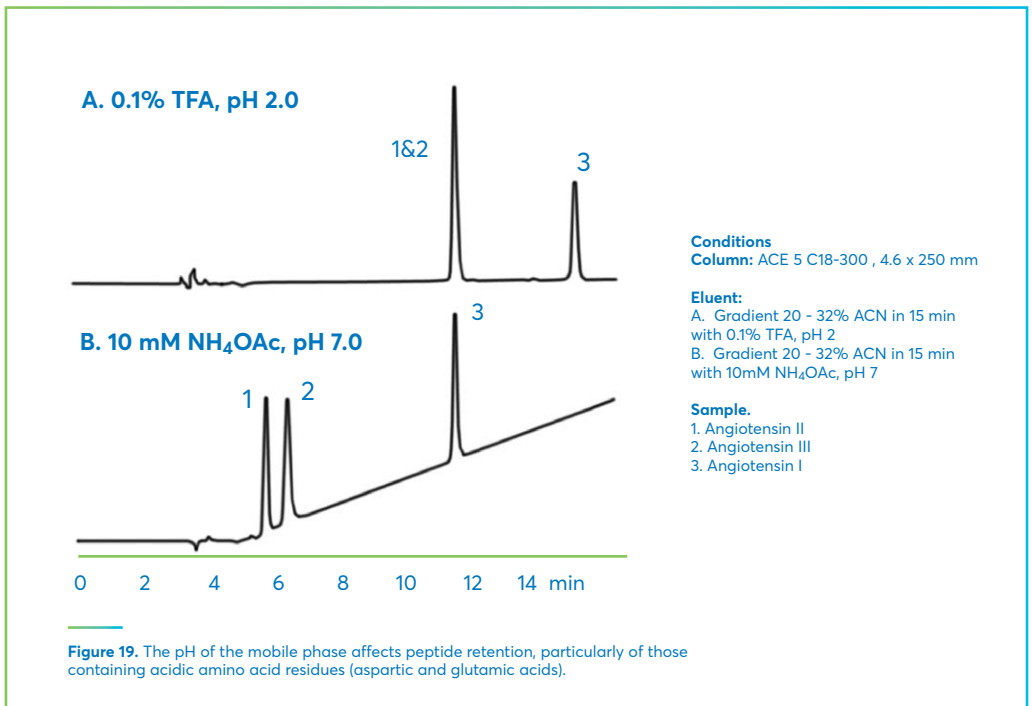
Heptafluorobutyric acid is sometimes used as an ion-pair reagent for basic proteins such as histones (Reference 9).





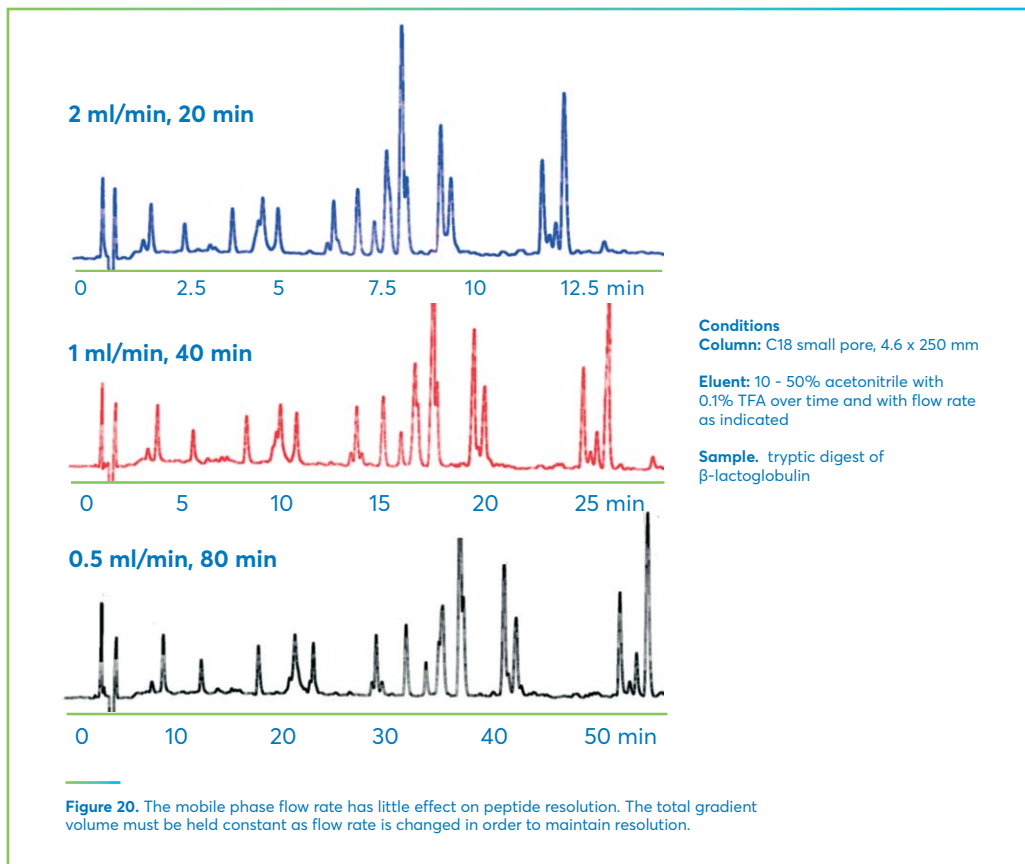
**The effect of pH on peptide retention.** Reversed-phase mobile phases used for peptide separations are generally adjusted to low pH, whether using TFA, phosphoric acid or other ion-pair reagents. At low pH, carboxylic acid groups - the carboxy terminal group and the side chains of aspartic and glutamic acid - will be protonated and only slightly polar. Increasing the mobile phase pH to 6 -7 will cause the carboxylic acid groups to ionize, making the peptide less hydrophobic. This reduces

the retention of all peptides, but particularly affects peptides containing aspartic or glutamic acid (Figure 19). The retention of aspartic and glutamic acid containing peptides will be reduced more than other peptides, thus changing selectivity. Increasing the mobile phase pH for peptide separations is becoming more frequent and, it can be useful in certain cases (see page 36 for an example).



**Flow rate.** The mobile phase flow rate has little effect on reversed-phase HPLC resolution. As illustrated in Figure 20, the resolution of peptides from a tryptic map is about the same when using flow rates of 0.5, 1.0 or 2.0 ml/min. The gradient volume however, must be constant in order to maintain the resolution. This requires reducing

the gradient time as the flow rate is increased. The system pressure increases as the flow rate is increased and pressure may limit the flow rate that can be used. Higher flow rates also decrease detection sensitivity slightly but may increase solubility of large or hydrophobic proteins.

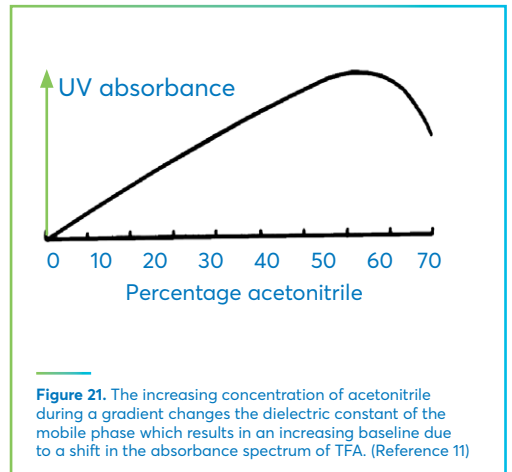
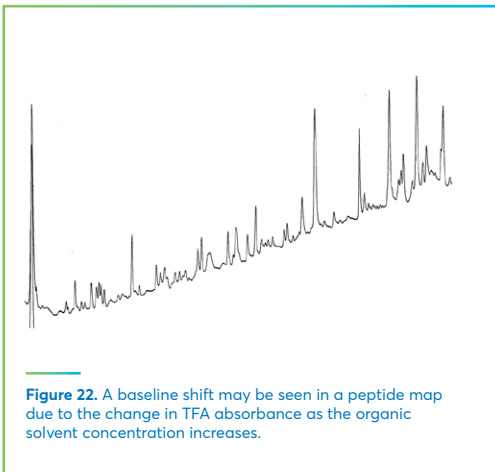


# Detection

When separating polypeptides by reversed-phase HPLC, detection is normally by UV absorption at 214-215 nm. The peptide bond absorbs well in this wavelength range and provides the most sensitive detection for polypeptides of all types.

One concern when using the low UV region for detection is absorbance by the mobile phase. Acetonitrile does not absorb UV light at 215 nm, however TFA absorbs slightly in this region. During gradient elution, the increasing concentration of organic solvent causes the dielectric constant of the solution to change which results in a shift in

the absorbance spectrum of TFA in the low UV region. This absorbance shift results in an upward drift of the baseline as illustrated in Figure 21. It is particularly noticeable when sensitive detection is required. As an example, peptide maps will often display an increasing baseline (Figure 22). A common practice to avoid this baseline shift is to reduce the TFA concentration in the organic solvent relative to the aqueous solvent. If 0.1% TFA is added to the aqueous solvent, then 0.08-0.09% TFA is added to the organic solvent. This results in a flatter baseline during peptide map elution.

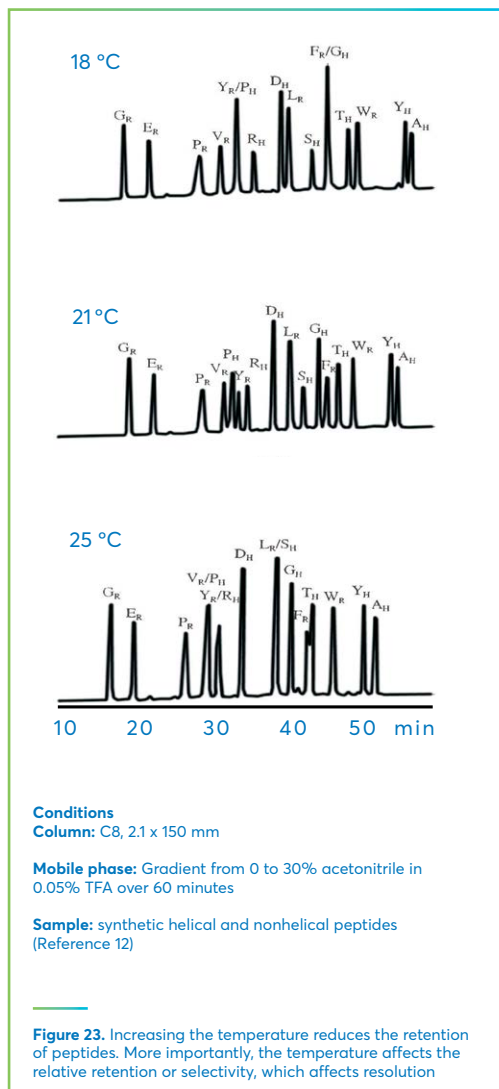


# Effect of temperature

Column and mobile phase temperature can affect peptide separations in two ways. Retention is reduced slightly with increased temperature. However, more importantly, the relative retention (selectivity) of peptide pairs depends on the temperature. Changes in selectivity with temperature affects resolution and temperature can be an important factor in optimizing the separation of a series of peptides, such as those produced during the protease digestion of a protein. The separation of a series of synthetic peptides as a function of temperature is shown in Figure 23. Peptide retention decreased slightly as the temperature was increased but, more importantly, resolution between peptide pairs changed with the changes in temperature.

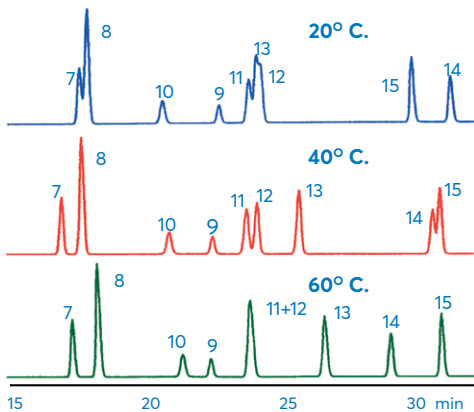
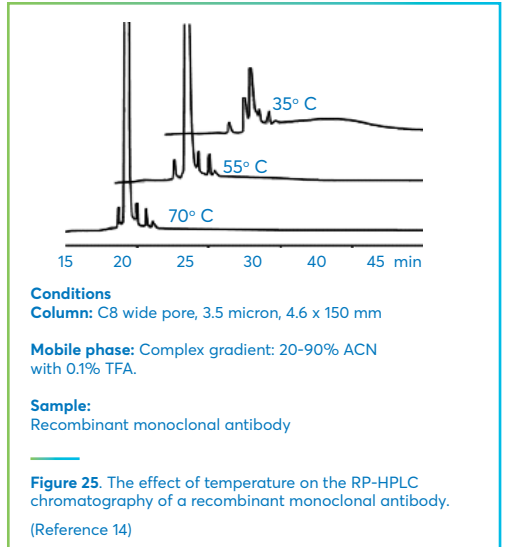
It is therefore important to control the temperature of column and mobile phase when separating peptides. Temperature is also an important variable to optimize when developing peptide separations, particularly separations of peptide maps.

The separation of peptides in a peptide map can be significantly affected by temperature as shown in Figure 24. In this example of the separation of peptides from the tryptic digest of human growth hormone, raising the temperature of the column not only affects peptide selectivity but can also cause a reversal on elution order. Peptides 7 and 8 are much better resolved at higher temperatures. Peptides 11 - 13 are best resolved at 40 °C and raising the temperature to 60 °C causes peptides 11 and 12 to nearly co-elute. Peptides 14 and 15 are well resolved at 20 °C with peptide 15 eluting first. Raising the temperature to 40 °C causes



peptide 14 to elute before peptide 15. Raising the temperature to 60 °C improves resolution between 14 and 15. Clearly, the column temperature can have a significant effect on resolution in a peptide map.

Protein resolution is less affected by temperature than peptide resolution, however protein peak shape and recovery is generally improved at higher temperatures, especially for large or hydrophobic proteins. Monoclonal antibodies chromatograph with poor peak shape at ambient temperatures but peak shape improves considerably when higher temperatures are used (Figure 25). Large or hydrophobic proteins often chromatograph best at elevated temperatures.



**Figure 24.** The relative retention of peptides from the digestion of a protein is often significantly affected by the temperature.

# Reversed-phase HPLC - MS

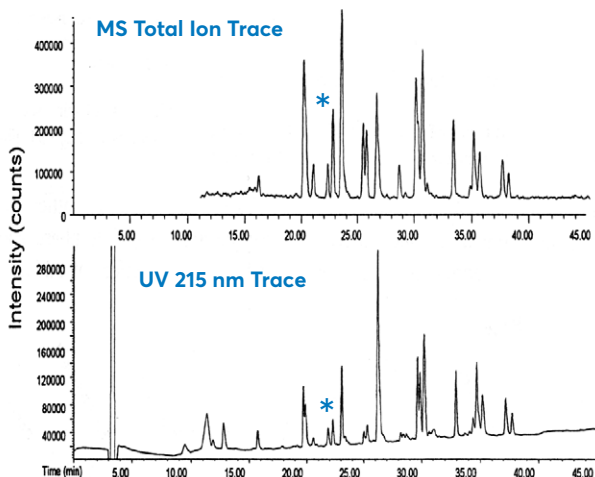
The combination of reversed-phase HPLC and mass spectrometry (MS) provides a powerful tool for protein/peptide analysis. Mass spectrometry interfaces with reversed-phase HPLC by means of the electrospray ion source, developed in the 1980s by John Fenn and colleagues. The benefits of using mass spectrometry on-line after chromatography include:

- MS is a very sensitive detection technique.
- MS provides the molecular weight of separated peptides/proteins.
- MS can specifically detect proteins or peptides by mass.
- Fragmentation information can help identify peptides.

- MS separates and measures based on charge and mass and is, therefore, "orthogonal" to reversed-phase, which separates by hydrophobicity.

HPLC-MS is widely used for the analysis of peptide maps, providing an orthogonal means of peptide detection (Reference 15). As illustrated in Figure 26, a total ion mass chromatogram parallels the UV chromatogram, however the peaks are of different magnitude because of orthogonal detection. Peaks that appear small by UV detection may, in fact, appear more significant by MS (Figure 26, see \*).

By measuring the molecular weight of the peptide(s) present in each of the peaks, mass

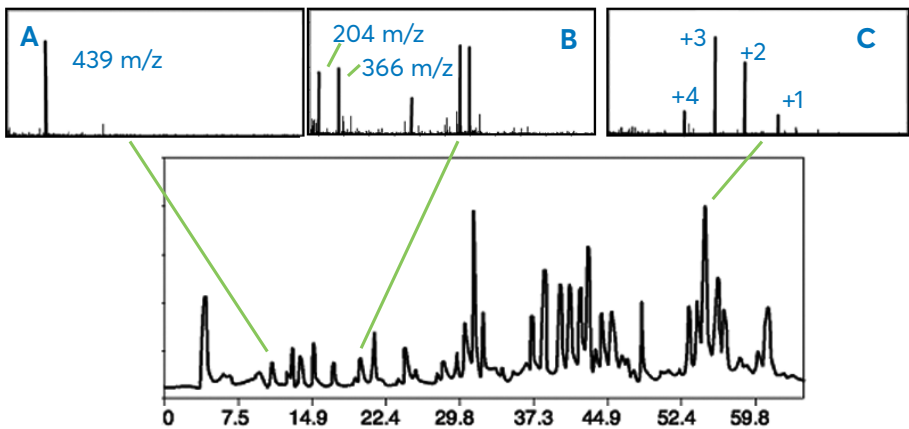


**Figure 26.** Separations of peptide maps can be monitored by both UV detection and mass spectrometry. Peak height is a function of the detection method and differs significantly between UV and MS traces. Especially note the pair of peptides marked by the \*. These appear to be small by UV detection but appear much more significant by MS. (Reference 16)

spectrometry provides useful information that assists in identifying the peaks separated by RP-HPLC. Since mass spectrometry is orthogonal to RP-HPLC, MS also confirms the purity of each peak or shows that two or more peptides are co-eluting and reports their molecular weights. Figure 27 illustrates this for three peaks in a peptide map. Peak A, which elutes early in the separation, is a 439 dalton tetrapeptide. Peak B is a glycopeptide, with an attached oligosaccharide or glycan. This is confirmed by the presence in the MS spectrum of Peak B of ions of  $m/z$  204 and 366, which are indicative of glycosylation. Peak C is a disulphide-linked dipeptide - two peptides connected by a disulphide bond. The peptide is shown to be

in four ions: +1, +2, +3 and +4. A +4 ion is only possible with dipeptides that have two amino termini and two basic amino acids. Two important aspects of HPLC-MS are the optimum flow rate of the electrospray interface and the effect of trifluoroacetic acid on peptide ionization.

The signal from a basic electrospray drops off rapidly above 5 - 10 microlitre/minute flow (Figure 28). This is incompatible with the flow rates used with standard analytical HPLC columns. Commercial electrosprays now provide a high shear flow of nitrogen gas parallel to the electrospray ("pneumatic-assisted electrospray")



**Figure 27.** Combining the separation power of reversed-phase with a second dimension of mass spectrometry provides considerable information on peptide maps. (Reference 18)

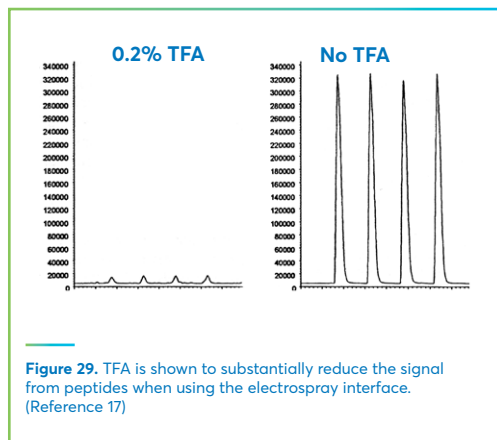
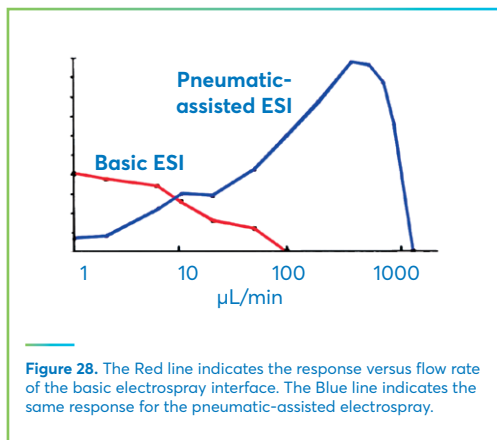
- A. This peak is identified as a small peptide by MS.
- B. The peak detected is identified as a glycan (sugar) containing peptide.
- C. This peak is a dipeptide, two peptides connected by a disulphide bond.

which boosts the optimal flow region of the electrospray to 200 - 500 microlitre/minute (Reference 19). This is still below the optimum flow rate normally used with standard analytical columns, consequently, scientists today generally use narrowbore columns (~2 mm i.d.) at flow rates of 200 - 300 microlitre/minute when using HPLC-MS.

As shown in Figure 29, the presence of TFA in the mobile phase flowing into the electrospray results in decreased signal for proteins and peptides. This is due to the very strong interaction between TFA and polypeptides, effectively neutralizing the polypeptide.

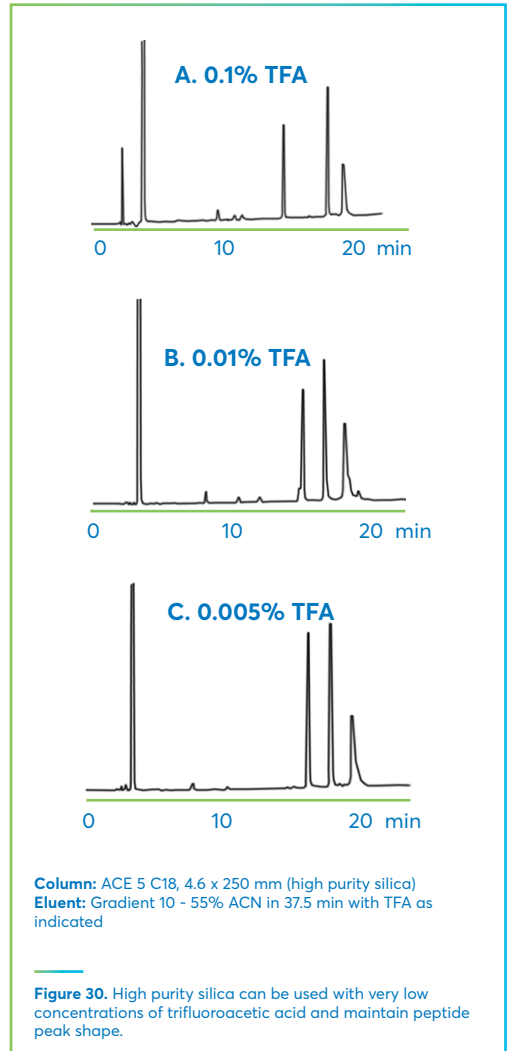
There are two approaches to correct the reduction of signal due to the TFA:

- Ignore the loss of signal. Often sufficient signal remains to obtain useful data even with the reduction in signal. In such cases the signal loss can be ignored.
- When the signal loss is too great, the best solution is to use high purity silica columns with reduced TFA concentrations. High purity columns can be used with very low concentrations of TFA and still maintain peak shape (Figure 30). Using high purity silica columns results in good signal response and good peak shape.





Other options include the use of formic acid in place of TFA or the replacement, post-column, of the TFA with acetic or propionic acid (Reference 17). Formic acid, however, does not give as good a peak shape as TFA or as good a separation, so performance is compromised. This is acceptable in proteomic applications but not in the analysis of protein therapeutic drugs. Replacement of the TFA post-column is cumbersome and causes the loss of resolution.



# Peptide mapping

Reversed-phase HPLC has become a standard method in the analysis and characterization of proteins, especially of therapeutic drug products. As a result of high resolving power and good detection sensitivity, reversed-phase analysis is able to provide a great deal of information about proteins. Proteins are sometimes analyzed as intact molecules, however, more often, the protein is broken in pieces by cleaving the backbone at specific amino acid residues using protease enzymes. The resulting peptide fragments are then analyzed by reversed-phase HPLC. This technique, called peptide mapping, is a standard approach to the analysis of proteins.

Much information about a protein can be obtained by reversed-phase analysis of the peptide fragments of the protein.

- Comparison of the peptide map of an expressed protein with that of a reference standard protein can ensure genetic purity and expression accuracy. Peptide maps are often used as an identity analysis tool for protein therapeutic drugs.

- Peptide maps allow the determination of protein degradation products such as deamidated asparagine and oxidized methionine.
- Peptide maps can determine or verify disulphide bond linkages which are required to ensure tertiary structure and therapeutic efficacy.
- Peptide maps can help determine the sites of glycosylation (addition of carbohydrate entities) as well as opening the door to detailed identification of the oligosaccharide attached.
- Peptide maps performed with mass spectrometry detection provide an advanced means of protein identification, peptide sequencing and data confirmation.

Protease digests are also used in the identification and quantitation of proteins in studying biological proteomes (Reference 20).

There are a number of protease enzymes which cleave the backbone of proteins, usually at specific amino acid residues. These include:

Protease	Specificity	Characteristics
Trypsin	Cleaves at the C-terminus of lysine and arginine	Produces on average one peptide for every 10-12 amino acids. Trypsin is the most commonly used protease.
Lys-C	Cleaves at the C-terminus of lysine	Produces 60-70% as many peptides from a protein as trypsin. The advantage of Lys-C is that it remains active in concentrations of urea as high as ~4 M.
S. aureus V8	Cleaves at C-terminus of acidic amino acids glutamic and aspartic acid.	Provides complementary information to trypsin.
Asp-N	Cleaves at the N-terminus of aspartic acid	Provides complementary information to trypsin.

Trypsin is the most commonly used enzyme to cleave proteins (protease). There are five steps in the digestion of proteins by trypsin: (**Note:** an excellent review of tryptic digestion is found in Reference 21)

- 01. Denaturation.** Proteins must be denatured to be digested in reasonable time. Proteins are denatured by incubation in a chaotropic reagent such as 6M guanidine HCl or 8M urea buffered at neutral pH (~ 7.5) for ~30 minutes at elevated temperature (37 °C).
- 02. Reduction of disulphide bonds.** Disulphide bonds prevent complete denaturation of a protein. The disulphide bonds in a protein being digested are usually reduced during the denaturation step by the addition of a reducing agent such as dithiothreitol (DTT) at a concentration of ~20 mM.
- 03. Carboxymethylation of free cysteines.** If the reduced cysteines remain free they are likely to

reform disulphide bonds which are incorrect.

To avoid this, free cysteines are methylated by the addition of a reagent such as iodoacetic acid at a concentration of ~ 60mM and incubated at 37 °C for thirty minutes. The reaction is quenched with 100 mM DTT.

- 04. Removal of salt.** Digestion cannot proceed in the presence of urea or guanidine because trypsin (a protein) will itself be denatured and lose enzyme activity. The urea or guanidine must be removed or reduced below 1M concentration by ion exchange or dialysis.
- 05. Trypsin digestion.** After salt removal the protein is dissolved in a buffer, either Tris or ammonium carbonate at pH 7.5 - 8.5 (the pH of maximal activity for trypsin), one part trypsin is added to 20-100 parts of the protein being digested and the protein is incubated at temperatures from subambient to 37 °C. Subambient temperatures require incubation times as long as sixteen hours. Digestion at

37 °C is complete in 1 - 4 hours, depending on the protein. If the time, temperature or relative concentration of trypsin is too low, digestion will be incomplete and potential cleavages may not occur, resulting in larger peptides with internal lysine or arginine. If the time, temperature or concentration of trypsin is too high then trypsin may digest itself, producing "autolysis" products, that is peptides from trypsin, confusing interpretation. A common practice is to perform a digestion omitting the protein but including the trypsin. The resulting sample is chromatographed under the same conditions as the full protein digest, thus revealing the extent of trypsin autolysis as well as the location of any trypsin peptides in the peptide map. Digestion time, temperature and relative concentrations of trypsin and protein are optimized in developing a tryptic digestion protocol. Recent advances in this area have included the introduction of digestion kits which can dramatically speed up this process.

When peptide maps are to be used to determine the location of disulphide bonds, a protein must be

digested without reducing the disulphide bonds. Many proteins, however, digest very slowly if the disulphide bonds are not reduced. In cases where digestion is slow or poor in the absence of a reducing agent, then Lys-C may be used in place of trypsin and the digestion performed in 4M urea to maintain the protein denatured during digestion. Surfactants that are sometimes used to maintain proteins in solution during digestion degrade chromatographic resolution and should be avoided.

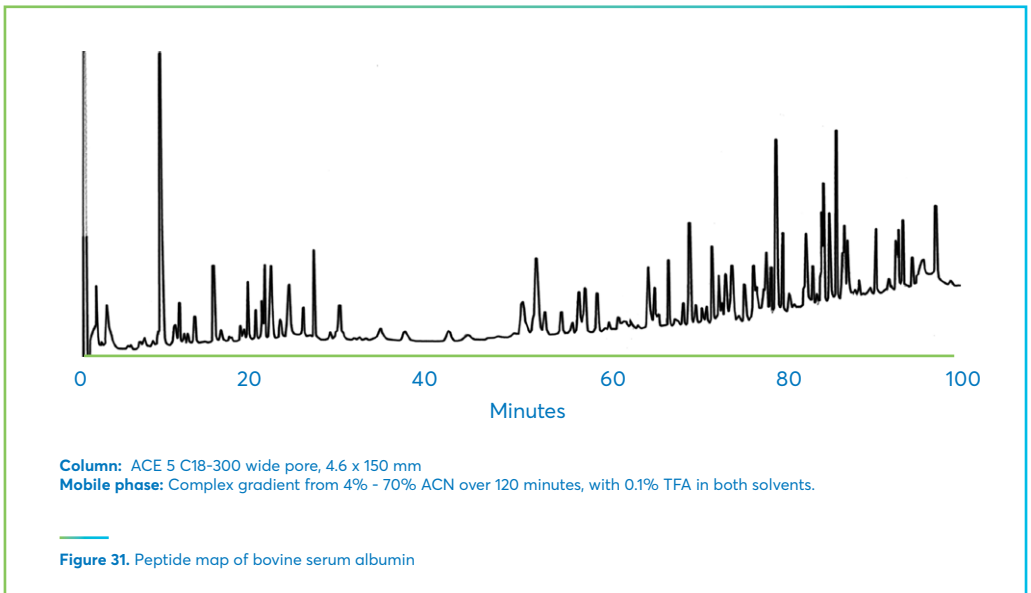
See Reference 22 for an example of a peptide map of a monoclonal antibody.

**Analysis of tryptic digestion.** The peptides from a protein digestion are analyzed by reversed-phase HPLC using a mobile phase containing TFA (see Pages 15 - 17) and with a gradient starting at about five percent acetonitrile (starting concentrations of acetonitrile of less than five percent may result in irreproducible chromatography of early eluting peptides) and increasing to 70 percent acetonitrile (See Figure 31). The length of the gradient time

depends on the size of the protein being digested. Large proteins produce more peptide fragments than small proteins and require longer gradient times to separate all of the fragments. The peptides from small proteins (less than ~20kd) can usually be separated in 45 - 60 minutes. Larger proteins (20-50kd) require longer gradient times of 60 - 120 minutes. The largest proteins (>50kd) require gradient times of 120 - 180 minutes. The flow rate is set at 1 - 2 ml/minute with the temperature adjusted to give the best resolution (See Pages 20 - 21). The column used is generally C18 reversed-phase. Columns with 100 angstrom

or 300 angstrom pores may be used and will often differ in selectivity. Long columns, and even column coupling can be employed to improve peak capacity.

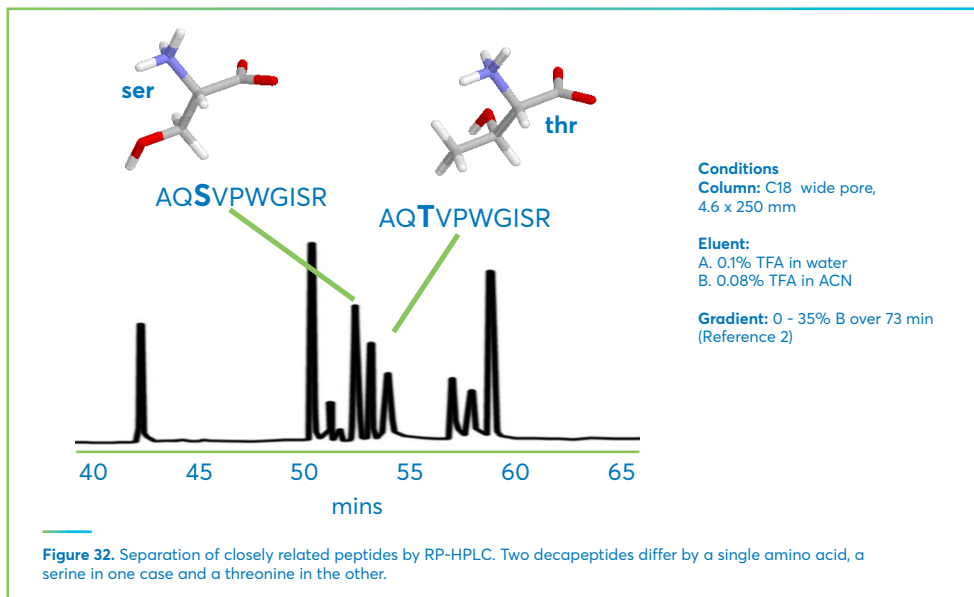
**Protein modifications result in changes in peptide retention.** If a protein is changed in any way due to translational or expression errors, degradation (deamidation, oxidation) or process variation, the change will be reflected in one or more of the peptide fragments. Due to the sensitivity of reversed-phase interactions with peptides, any changes in a peptide result in a change in the



retention time of that peptide. In the example in Figure 32, two decapeptides that differ by one amino acid - a threonine in one and a serine in the other - are separated by reversed-phase HPLC. Not only is the difference only a single amino acid, but both of the amino acids are hydroxy amino acids and they differ by the addition of a methyl group on the threonine side chain. This illustrates that any change to a protein will be reflected in a change in a peptide which leads to a shift in the retention of that peptide. The essence of peptide mapping is the ability of reversed-phase HPLC to resolve peptides with minor differences.

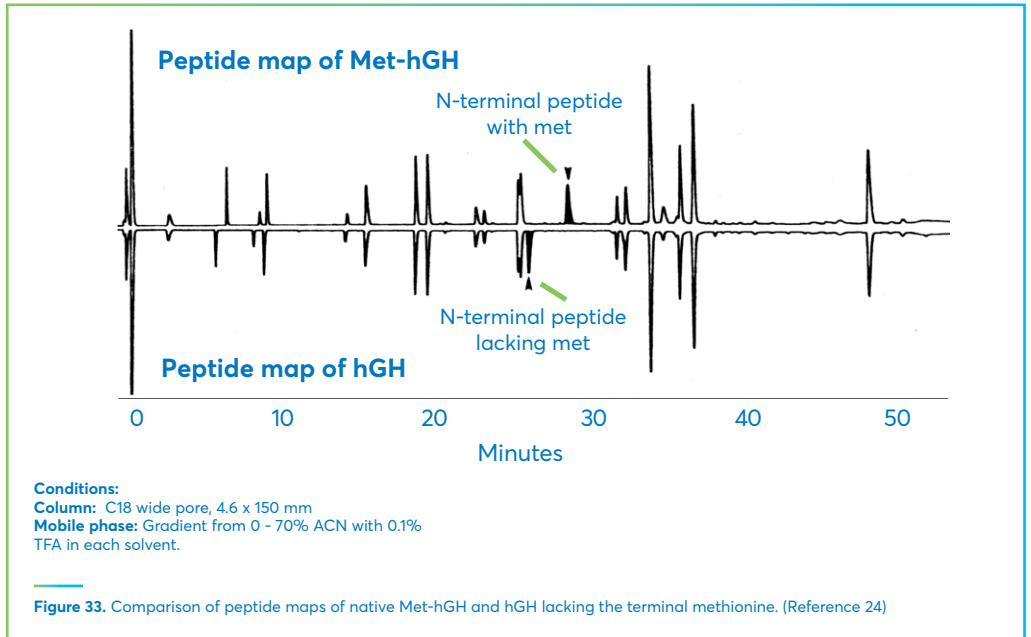
**Comparison of the peptide map of test sample with the peptide map of a reference protein.**

Peptide maps provide much information about proteins. A common practice is to compare the peptide map of a test protein with the peptide map of a reference protein. In Figure 33, the peptide map of methionyl human growth hormone (as expressed in E. Coli) is compared with the peptide map of native human growth hormone (lacking the methionine). The meth-hGH peptide elutes later than the non-met hGH peptide because methionine is hydrophobic. In this example the second map is inverted for ease



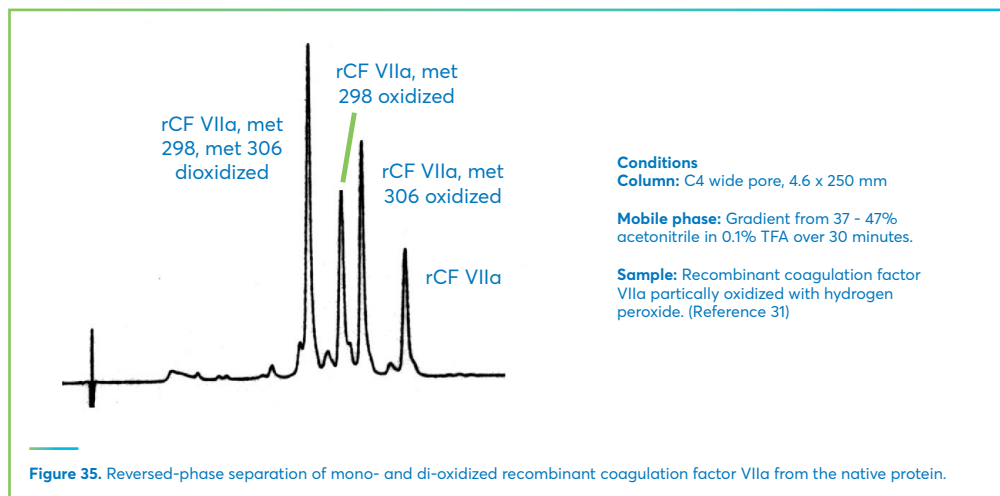
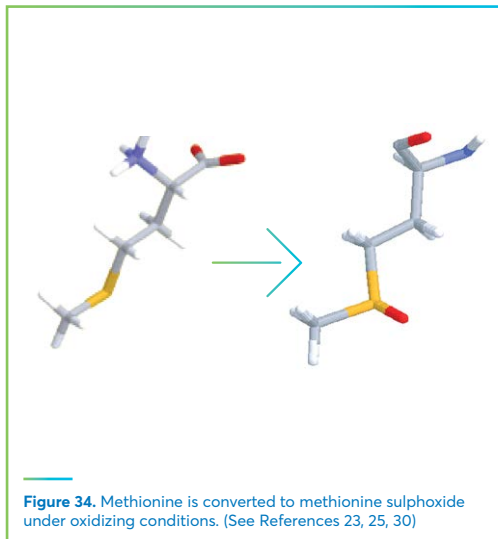
of comparison. In some cases a mixture of the two digestion mixtures is chromatographed in order to confirm minor changes and ensure that analytical artifacts are not responsible for changes in the peptide map.

Peptide map comparisons reveal changes and modifications to proteins such as genetic changes, mis-translations, protein degradations (deamidation, oxidation) and changes in post-translational modifications.



# Protein oxidation

When a protein experiences oxidizing conditions, although several amino acids may be affected, the most likely amino acid to be oxidized is methionine, which is converted to methionine sulphoxide (Figure 34). The oxidation of a methionine residue depends on its location in the protein. A methionine buried in the interior of a protein is not likely to be oxidized. Methionine side chains near the surface and in contact with the solvent are most likely to be oxidized. Oxidizing conditions include heat, the presence of transition metals and the presence of oxygen in the solution. Like deamidation, methionine oxidation may lead to a loss or reduction of biological activity or, in some cases, have no effect on biological activity. This depends on the location of the methionine in the protein. However, the presence of a methionine sulphoxide indicates that the protein has experienced oxidizing conditions and,



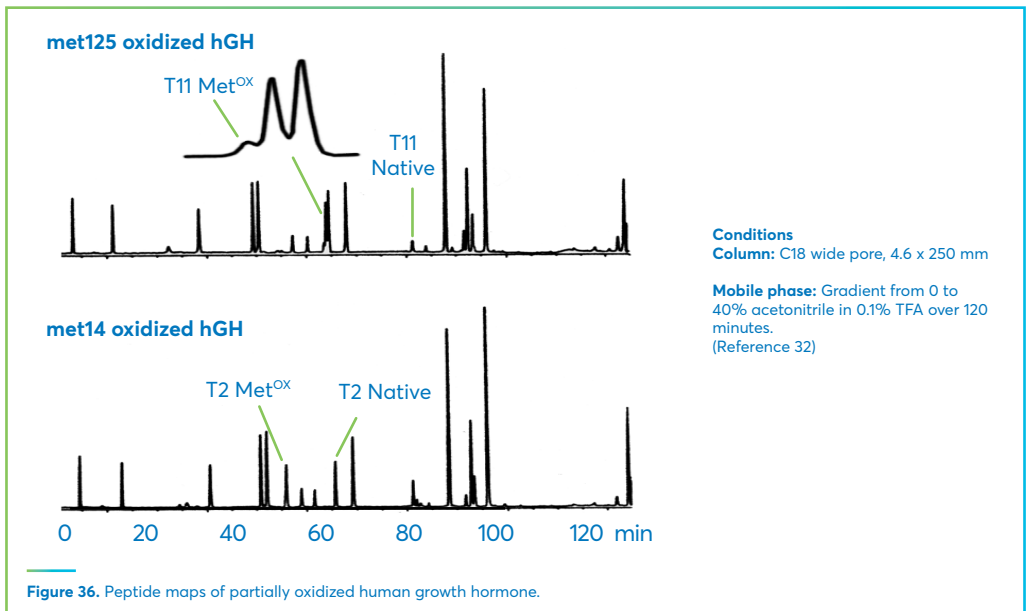


therefore, methionine oxidation is an indicator of the protein having experienced stress conditions.

Protein therapeutic drugs are routinely monitored for oxidation both because of the possible effects on biological activity and as an indicator. The oxidation of a methionine causes a decrease in the hydrophobicity of a protein and consequently an oxidized protein will elute before the native protein in reversed-phase HPLC as illustrated in Figure 35. Recombinant coagulation factor VIIa (rCF VIIa) was partially oxidized with hydrogen peroxide and the solution analyzed by reversed-phase HPLC. The dioxidized protein (two oxidized methionine residues) elutes first. The

two variants with a single oxidized methionine elute next followed by the native protein. Good resolution was seen between all four species. The remarkable separation is that between the two singly oxidized protein variants. These each have one oxidized methionine and differ only in which methionine is oxidized.

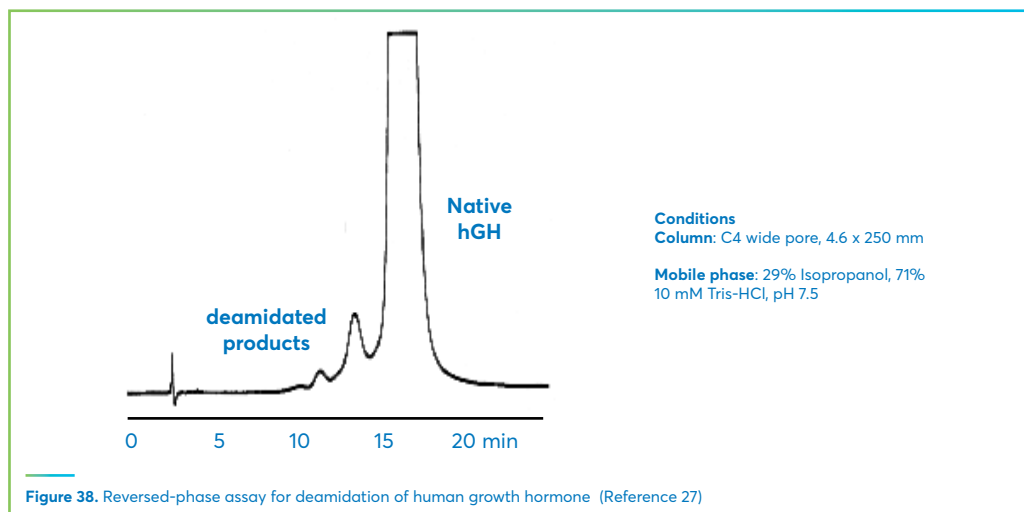
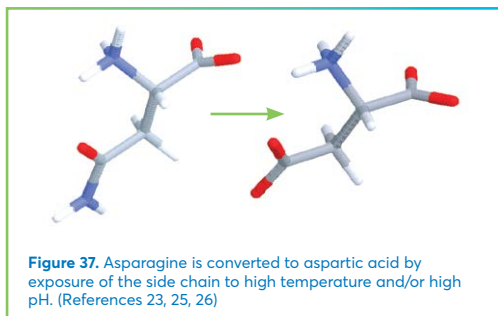
Methionine oxidation is usually monitored using a peptide map. As shown in Figure 36, peptides containing oxidized methionine elute significantly earlier than peptides with native methionine. T2 met<sup>ox</sup> elutes well before the native T2 tryptic fragment and T11 met<sup>ox</sup> elutes well before T11 native, making identification and monitoring easy.



# Protein deamidation

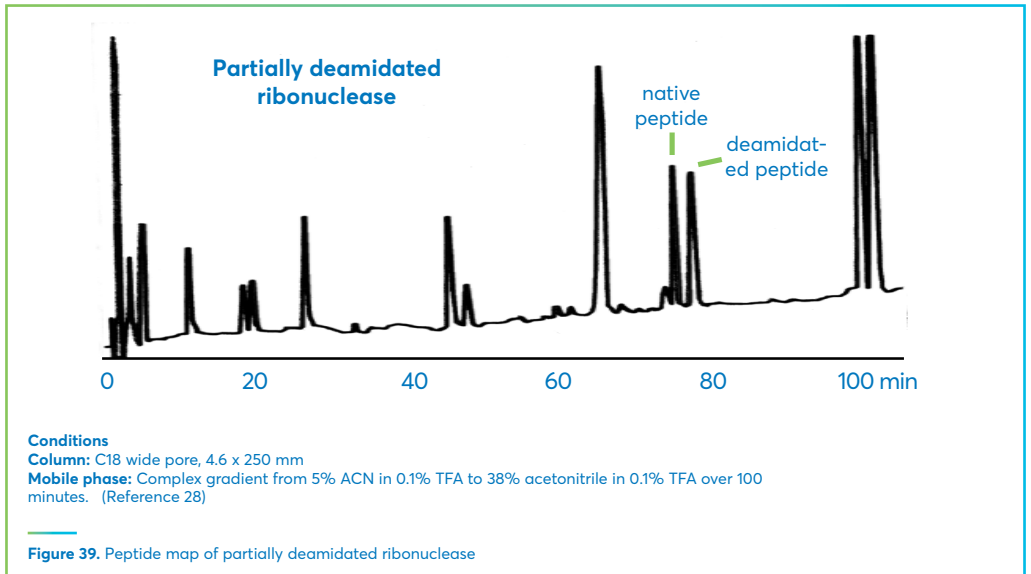
When subjected to either high temperatures or high pH, proteins often degrade. The most likely chemical degradation to occur under stressed conditions is the conversion of the amide side chain of asparagine to either aspartic acid or isoaspartic acid (Figure 37). Many proteins lose biological activity when an asparagine is deamidated although the biological activity of some proteins is not affected. Even when deamidation does not cause a reduction of biological activity, deamidation acts as an indicator of protein exposure to adverse conditions. The likelihood of deamidation of a specific asparagine residue depends on its location within the tertiary structure of the protein. The only asparagine residues that will deamidate are those near the surface with access to the solvent. Adjacent amino acid residues also affect

the rate of deamidation. Glycine located next to the asparagine greatly increases the likelihood of deamidation. Large hydrophobic amino acids such as leucine and isoleucine next to the asparagine reduce the rate of deamidation.



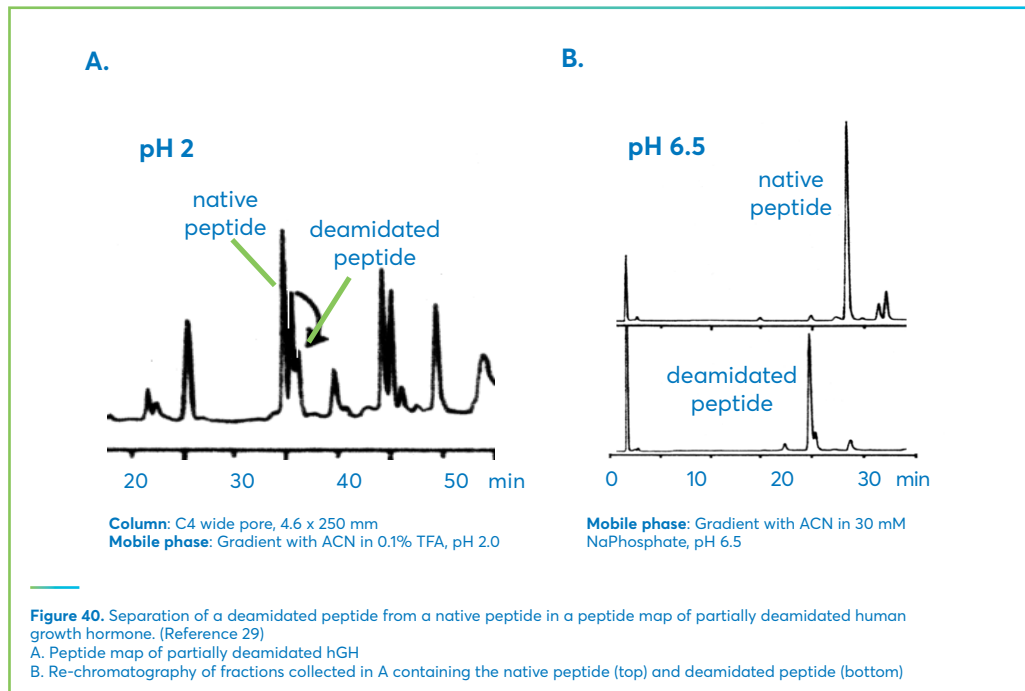
Since deamidation may affect biological activity and indicates exposure to adverse conditions, it is common practice to monitor protein therapeutic drugs for asparagine deamidation. This is commonly done using reversed-phase HPLC. An assay measuring deamidation of human growth hormone by the reversed-phase chromatography of the intact protein molecule is shown in Figure 38. This assay is run at pH 7.5 so that the aspartic acid resulting from deamidation is ionized. This makes the deamidated growth hormone less hydrophobic than the native protein, causing it to elute before the native protein.

A more common approach for determining deamidation is by the retention of asparagine/aspartic acid containing peptides in a peptide map. In peptide maps run at pH ~2, a deamidated peptide containing aspartic acid will be slightly more hydrophobic than the native asparagine-containing peptide and thus elutes slightly later than the native peptide as illustrated in Figure 39. The deamidated peptide is easily identified and measured, providing a good method for measurement of deamidation.



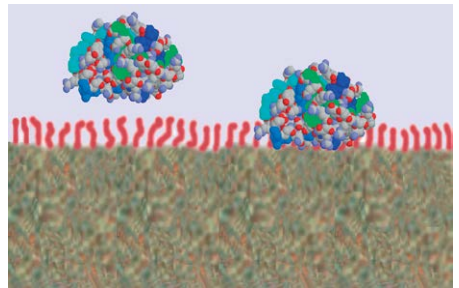
Sometimes a deamidated peptide is either poorly separated from the native peptide or coelutes with another peptide in the map. If resolution is insufficient, the peptide map can be run at elevated pH. As shown in Figure 40, a deamidated peptide that is poorly resolved at low pH can be well separated from the native peptide by using a pH of 6.5. In the low pH peptide map of partially deamidated human growth hormone (Figure

40A) the deamidated peptide elutes slightly later than the native peptide but is not resolved from another peptide in the map. The peptide peaks were collected and re-chromatographed at higher pH (pH 6.5). The aspartic acid in the deamidated peptide is ionized at higher pH and the deamidated peptide elutes well before the native, asparagine-containing peptide (Figure 40B).

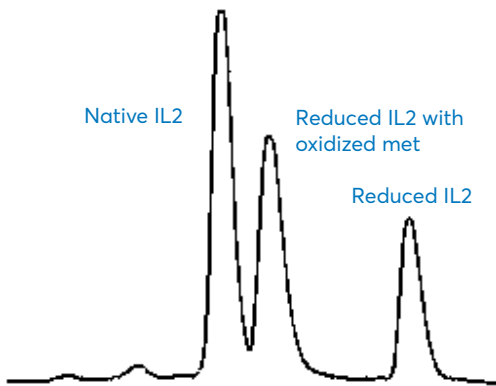


# Disulphide bond determination

Proteins rely on correct disulphide bonding to maintain tertiary structure and biological activity. If disulphide bonds are reduced or switched, a protein will lose native tertiary structure and will not be biologically active. HPLC retention depends on the size of a protein's "hydrophobic foot" (Figure 41), which is affected by the tertiary structure. Changes in disulphide bonding usually increase the size of the "hydrophobic foot", which causes an increase in reversed-phase HPLC retention of the protein. In Figure 42, native interleukin II eluted significantly earlier than reduced interleukin II due to the change in tertiary structure when the disulphide bonds were reduced.



**Figure 41.** Protein retention depends on the size of the "hydrophobic foot". Reduced disulphide bonds allow the protein structure to partially denature, which enlarges the "hydrophobic foot" and increases reversed-phase retention.



#### Conditions

**Column:** C18 wide pore, 4.6 x 250 mm

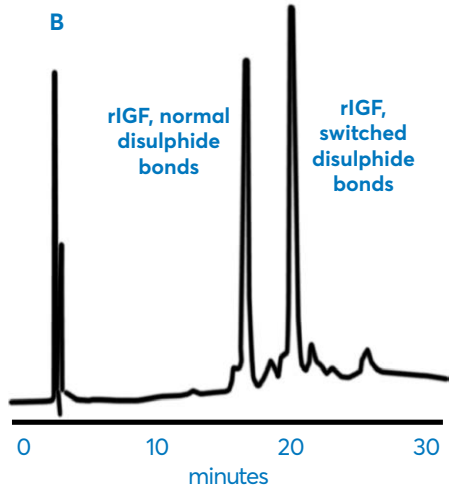
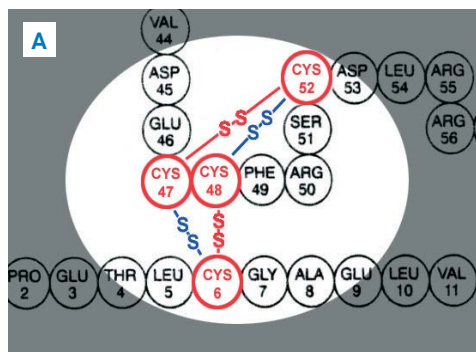
**Mobile phase:** Gradient from 44.5 to 50.8% ACN over 90 minutes at 2 ml/min

**Sample:** Interleukin II muteins (Reference 33)

**Figure 42.** The "hydrophobic foot" of interleukin II enlarges with the reduction of disulphide bonds, resulting in increased retention in reversed-phase HPLC.

Even small changes in disulphide bonding can have sufficient effect on the hydrophobic foot to change retention. Native insulin-like growth factor (IGF) has disulphide bonds between Cys52 and Cys47 and between Cys48 and Cys6 (Figure 43A RED). After being subjected to 36 hours of air oxidation, the disulphide bonds switched to between Cys52 and Cys48 and between Cys47 and

Cys6 (Figure 43A BLUE) in some of the IGF protein molecules. Insulin-like growth factor with the switched disulphide bonds elutes later than the native protein (Figure 43), indicating an increase in the size of the hydrophobic foot. Reversed-phase chromatography of a native protein will often reveal changes in disulphide bonding or tertiary structure by a change in reversed-phase retention.

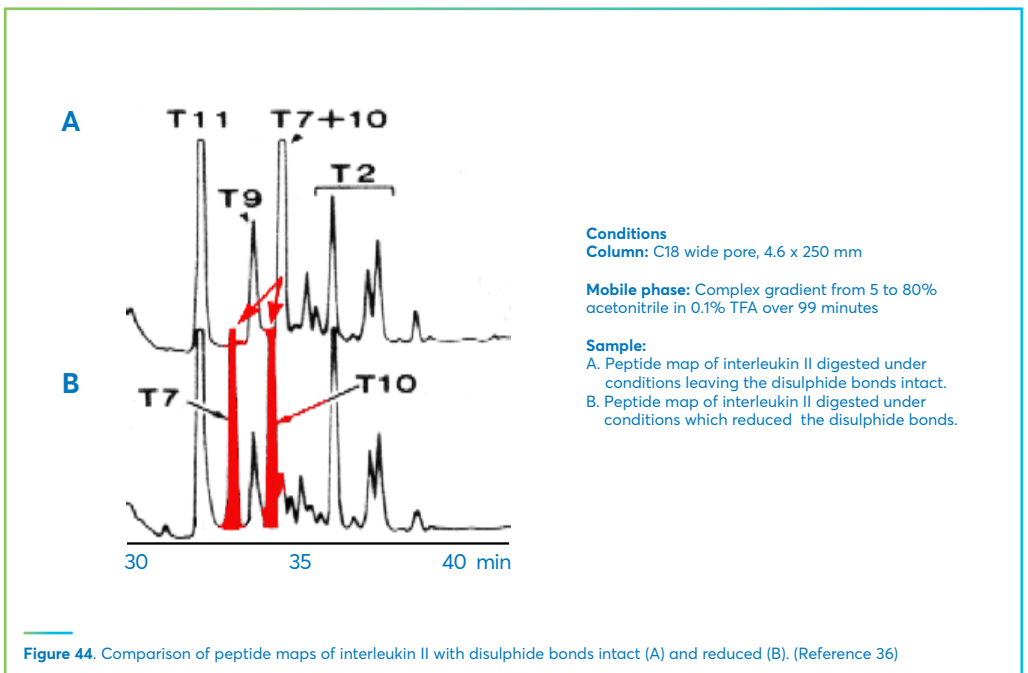


**Conditions**  
**Column:** C4 wide pore, 4.6 x 250 mm  
**Mobile phase:** Gradient from 20 to 38% acetonitrile:isopropanol (88:12) in 0.1% TFA over 27 minutes. (Reference 34)

**Figure 43.** Separation of insulin-like growth factor with normal disulphide linkages and with switched disulphide bonds. (Reference 26)

If the disulphide bonds are not reduced during the protease digestion of a protein (by omitting the DTT and carboxymethylation steps), then the peptide map will contain disulphide linked dipeptides. Comparing a non-reduced map with a normal peptide map with reduced disulphide bonds, allows determination of the location of the disulphide bonds. Interleukin II was digested under both reducing and non-reducing conditions and the peptide maps of both digestions were run by RP-HPLC (Figure 44). In Figure 44A the peptide labelled T7+T10 is a disulphide linked dipeptide.

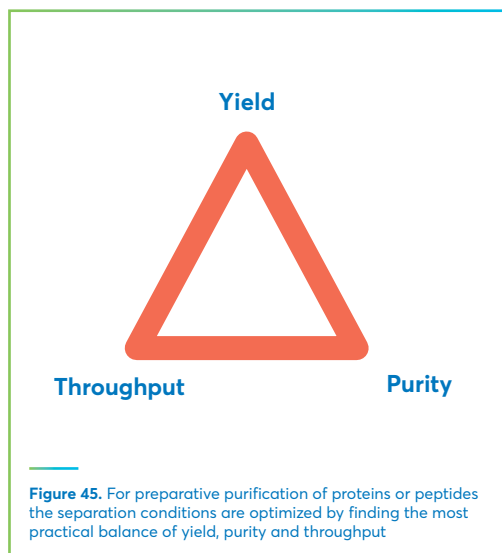
The peptide map produced with disulphide reduction (Figure 44B) reveals two peptides which appear as T7 and T10. This confirms that the disulphide bond is between these two peptides. Monitoring all of the peptides in the protein digest allows assignment and/or confirmation of all of the disulphide bonds in the protein. If two cysteines are present in a tryptic peptide, then an alternative protease must be used to determine which cysteine is involved in the disulphide bond. Peptide mapping coupled with mass spectrometry is often used to determine the location and state of disulphide bonds (Reference 35).



# Protein/peptide purification

Reversed-phase HPLC is an effective tool for the purification of proteins and peptides. Target proteins/peptides can often be resolved from impurities by reversed-phase HPLC, collected as fractions and the fractions used for further studies, for analysis by orthogonal analytical techniques or even for use as a therapeutic drug.

When analyzing proteins/peptides the goal in developing chromatographic conditions is to optimize **RESOLUTION** and **TIME**. When separating proteins/peptides with preparative chromatography, developing chromatographic conditions involves optimizing three parameters (see Figure 45):



- **Yield** is the amount of purified target protein/peptide that is obtained from each chromatography step. Higher yields result in more practical purification processes and lower costs.
- **Purity** indicates the extent to which impurities have been removed from the desired product. Higher purity results in better data from subsequent analysis or a product with few impurities.
- **Throughput** is a measure of how much material is purified in a preparative cycle. High throughput means more material for study or analysis or more drug substance for pharmaceutical use at a given cost and time.

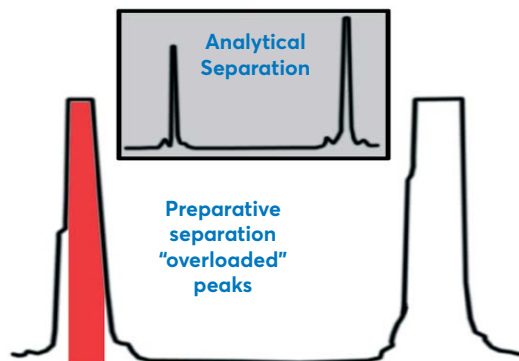
Since the goals of preparative purification differ from those of analytical chromatography, separation conditions are optimized differently.



**Sample loading.** In analytical chromatography, small samples are loaded onto the column to ensure that the amount of sample loaded does not affect resolution. If the amount of sample loaded is too high, peaks will become broader and resolution will be reduced. The amount of sample that can be loaded onto the column without band broadening occurring (the "sample capacity") depends on the size of the column (See Appendix for a list of column dimensions and sample capacity). When purifying a protein/peptide preparatively it is common practice to exceed the sample capacity, that is to "overload" the column, in order to increase yield and throughput (Figure 46). As much as 10 - 50 times the sample capacity can be loaded onto a column if some loss of resolution is acceptable (Appendix, Maximum Practical Load). Although the peaks in Figure 46 are broadened due to

column overload, peak shape remains fairly good, indicating that significant overload can increase yield while preserving purity, although some loss of the target protein/peptide is unavoidable. The peaks in Figure 47 are similarly broadened due to sample overload.

**Fraction collection, analysis and pooling.** When a column is overloaded, it is common to discard the beginning and end of the peak. In Figure 46 the centre of the peak indicated by the red area is collected. The leading and trailing edges are discarded. This avoids collecting poorly resolved impurities and results in increased purity and decreased yield. In preparative chromatography several fractions are collected across the peak of interest and these are analyzed for impurities. Based on the analytical results, fractions with little or no impurities are pooled together while those



**Figure 46.** In an example of peptide purification, preparative separation involves overloading the column with sample in order to increase yield. Resolution is reduced and the beginning and end of a peak must be discarded in order to optimize purity, even though yield is slightly reduced.

fractions near the beginning and end of a peak which contain significant impurities are discarded. Choosing which fractions to pool and which to discard is a matter of balancing purity and yield.

For instance, an “analytical” column of 4.6 x 250 mm dimensions can be used to purify small amounts of polypeptide, up to about 200 micrograms, without loss of resolution. However, in order to improve yield and throughput as much as 10 milligrams can be purified on a column of these dimensions with some loss of purity or with some loss of yield. Careful selection of the portion of the peak to collect results in an optimum balance between purity and yield.

Although most of the focus in preparative chromatography is on sample mass, sample volumes may also be large in preparative chromatography. Although sample loops and injectors may be used, much higher volumes of sample can be injected by “pumping” the sample onto the column. The inlet line of the pump is placed in the sample container and the sample is loaded onto the column by the elution pump itself. Very large volumes of sample can be loaded in this manner if loaded at low organic solvent concentration (generally the sample is loaded in an aqueous media) and the target protein/peptide is eluted with a gradient of increasing organic solvent.

**Adsorbent particle size.** While five micron particles are most commonly used for analytical chromatography, preparative chromatography often uses adsorbents with larger particle sizes. Particularly when loading the column

with a sample in excess of the sample capacity (overloading the column) column efficiency is of less importance than in analytical chromatography. Columns packed with large particles separate proteins and peptides as well as columns packed with small particles if the column is overloaded. Consequently adsorbents of 10 micron and higher are often used for preparative chromatography. The distribution of particle sizes is often greater as well. Rather than very narrow particle size ranges of 0.5 micron or less, for preparative chromatography larger particle size ranges such as 10 - 15 micron are common. Larger particles are preferred for preparative chromatography because of reduced back pressure and lower cost.

**Column diameter.** Small bore columns (less than 2 mm diameter) are rarely used for purification purposes because of very low sample capacity (Appendix). Small scale laboratory purification may be accomplished on narrow bore (~2 mm diameter) and analytical (4.6 mm diameter) columns. The chromatography conditions used for such small scale preparative separations are typically the same as those used for analytical separations. Where larger amounts of protein/peptide are needed, columns of 10 mm and 22 mm diameter are used. As much as 1 mg of protein or peptide can be purified on a 10 mm column and 5 mg can be purified on a 22 mm column. Even more protein/peptide can be purified if column overload is acceptable, up to 50 mg on a 10 mm column and 200 mg on a 22 mm column. For purification of larger amounts of protein or peptide, larger diameter columns of 50 mm, 100 mm or even greater are used. As much as 5 grams

have been reported to be purified on columns of 50 mm diameter.

**Column Length.** Preparative columns tend to be relatively short compared to analytical columns. This is because total column volume is more important than length in preparative chromatography, particularly for proteins. Columns with diameters as large as 60 cm and lengths of 12-15 cm ("pancake" columns) have been used for large scale purification of protein therapeutic drugs (Reference 37). Column dimensions are optimized for practicality rather than efficiency because columns are generally overloaded in preparative chromatography and efficiency is much less important than yield, purity and throughput.

**Mobile phase components.** As in analytical chromatography, acetonitrile containing TFA is commonly used for small scale purification using columns up to 10 or 22 mm diameter. Larger scale purification often employs solvents such as ethanol in place of the acetonitrile and acetic acid in place of the TFA. While resolution is reduced using these mobile phases, they are more practical for large scale use and the loss in resolution is parallel with the loss of resolution inherent in overloading of columns.

References 38, 39 and 40 describe general conditions and sample loading for preparative chromatography of proteins and peptides.

**Protein denaturation.** It is commonly thought

that reversed-phase chromatography denatures proteins and thus the eluting protein does not represent the native protein and may not be biologically active. Although reversed-phase HPLC does occur under conditions that can denature proteins, native, biologically active proteins can be obtained after reversed-phase chromatography. The organic solvent may weaken hydrophobic forces, causing some loss of the tertiary structure of the protein. The hydrophobic surface of the adsorbent also tends to cause some unfolding of a protein. Protein unfolding, however, is usually slow compared with chromatographic times and proteins may only be slightly denatured during reversed-phase chromatography. Since disulphide bonds continue to hold the protein in a globular form and unfolding is only partial, a protein eluting from a reversed-phase column can often be restored to its native form by incubating the eluted protein in an appropriate refolding buffer and allowing the native structure to reform. There are many examples which show that reversed-phase HPLC is capable of producing purified protein with native tertiary structure and biological activity. Reference 41 describes the reversed-phase purification of trypsin, which retains activity and is subsequently used for tryptic digestions of proteins. Recombinant erythropoietin, a commercially successful protein therapeutic, employs reversed-phase HPLC as part of the isolation of the protein drug from its cell culture expression system (Reference 42). Granulocyte stimulating factor, another commercial protein therapeutic, is purified using reversed-phase HPLC

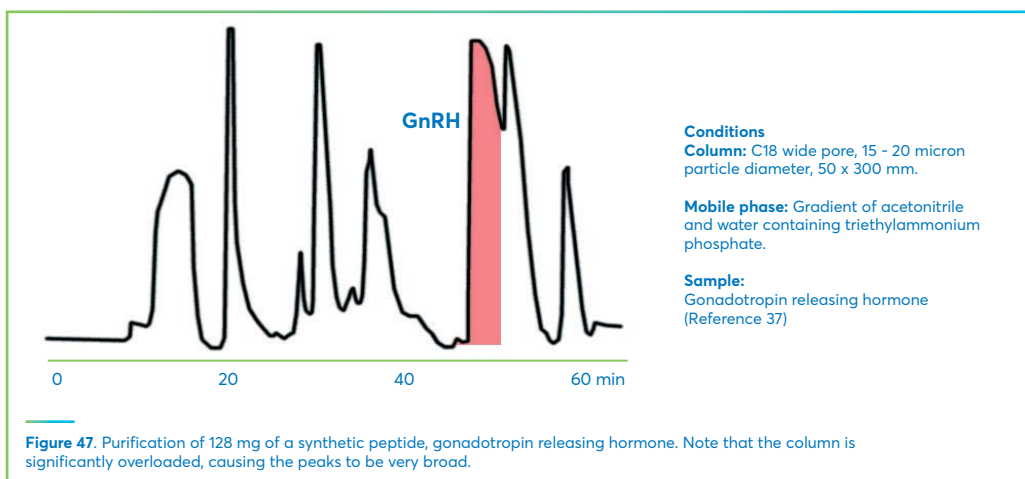
(References 43, 44). Recombinant human insulin has also been purified in active form using reversed-phase HPLC (Reference 45).

**Purification example.** Figure 47 shows an example of the purification of a synthetic peptide, gonadotropin releasing hormone (GnRH) antagonist. This purification procedure was developed by:

- Establishing elution conditions on an analytical (4.6 x 250 mm) column.
- Loading 1.2 grams of a synthetic peptide mixture on a 50 x 300 mm column and eluting based on the conditions established in the first step (Figure 47)
- Fractions collected across the GnRH antagonist peak were collected and analyzed by the analytical method and pooled for maximum yield and purity.

- Fractions were desalted by re-chromatography on a reversed-phase column using acetonitrile and trifluoroacetic acid as the eluting solvent.
- Final fractions were pooled for maximum yield and purity

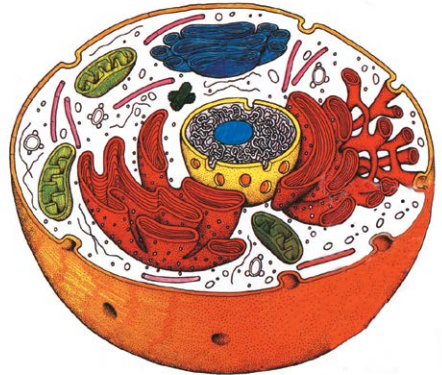
From this single chromatographic purification step, 128 mg of purified peptide were collected from a reaction mixture of 1.2 gm. The purification used the same organic solvent used in analytical separations, acetonitrile, but used triethylammonium phosphate in place of trifluoroacetic acid. The peaks are broadened due to the column overload and the chromatogram is not as well resolved as analytical chromatograms. However the peaks remain fairly compact and the eluent containing the target peptide, GnRH, is easily collected.



# Proteomics

**Proteomics** is the identification and quantitation of cellular proteins in order to understand biological changes, disease states and to develop biomarkers for diseases and targets for therapeutic drugs.

Mammalian cells contain as many as 30,000 - 40,000 proteins if one counts modified versions of one basic protein as one protein. If the variously modified proteins are each counted individually, the number of proteins is well over 100,000. The abundance or concentration of different proteins in a cell can vary by several orders of magnitude. Any change to a cellular system, such as ageing, becoming diseased or being treated with a therapeutic drug, will cause a change in the relative abundance or expression of one or more of these proteins. Proteomics seeks to identify each protein in a cellular system and to identify and monitor changes to the protein such as modifications to the protein itself (post-translational modifications) or changes in its relative concentration. If a certain event causes a protein abundance to increase we say that the protein is **up-regulated** by the event. If an event causes the protein abundance to decrease we say that the protein has been **down-regulated**. Protein abundances may be changed significantly when affected by cellular changes and changes in protein abundance are signals of the changing event.



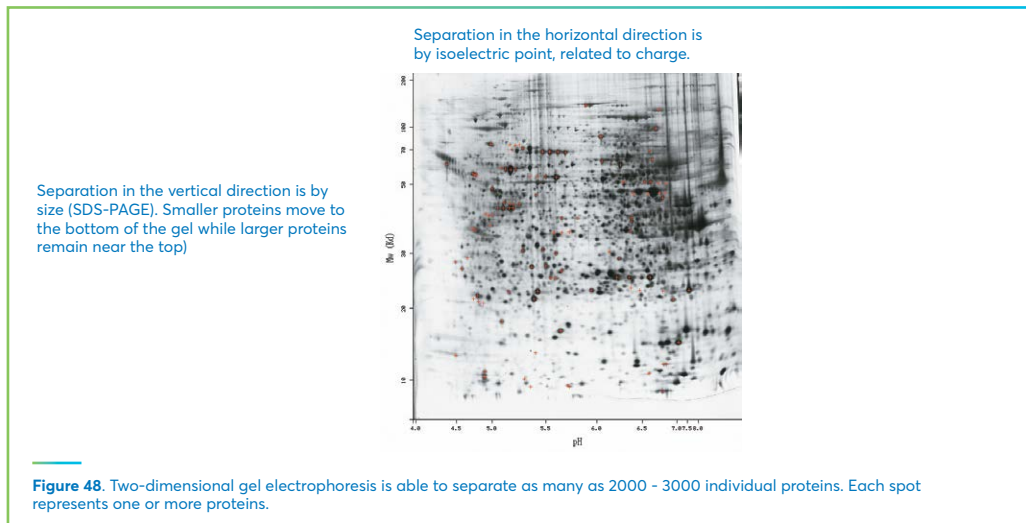
## Post-translation modifications (PTMs)

are changes to a protein after it has been synthesized by the cell. PTMs include the addition of oligosaccharide chains (glycosylation), modification of the N-terminus of a protein with an alkyl group such as acetate (acetylation, etc), attachment of a phosphate group to a serine, threonine or tyrosine (phosphorylation) and similar events. PTMs such as phosphorylation are largely temporary and are used by the cell to transmit messages within the cell and through the cell walls. PTMs such as glycosylation are structural and often affect protein folding, interactions with other molecules and interaction with cell walls. All of these represent the dynamics which occur within a living cell.

**A proteome** represents all of the proteins present in a cell or in any sample being studied. A complete cellular proteome can consist of as many as 40,000 individual proteins with up to 10 - 20 different modified forms and varying in abundance by several orders of magnitude. The goals of proteomics are clearly enormous, requiring the most powerful of separation and analysis techniques.

**Electrophoresis.** In order to identify and/or quantitate proteins, individual proteins must be separated from other proteins as much as possible. For a great many years the primary means of separating intact proteins was by gel electrophoresis. Two dimensional gel electrophoresis (2DGE) separates proteins first

by their isoelectric point and then by size (see Figure 48). 2DGE of entire cell proteomes is extremely complex and, while separating as many as 2000 - 3000 proteins, many proteins are either not separated from each other or are not visible on the gel because of low abundance. 2DGE remains a powerful tool for protein separation. It is a well developed technique with many experienced users, is capable of high resolution and can provide quantitation of many proteins. Its limitations are that it requires a lot of manual effort and is time consuming, it primarily measures more abundant proteins and has difficulty measuring some important proteins such as membrane bound proteins.



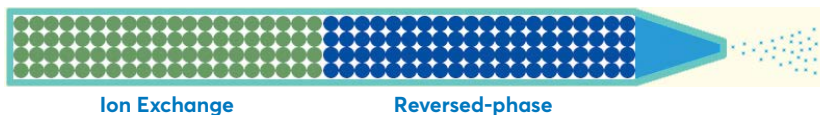
**Chromatography.** Chromatographic techniques have developed into powerful separation techniques, capable of separating large numbers of proteins and peptides. Any given chromatography technique by itself, however, will still be able to separate only a small fraction of cellular proteins. As a result, combining chromatography techniques has become a widespread method for protein separations for proteomic analysis.

#### Two dimensional chromatography (2D-LC).

Following a pattern long used in protein purification, John Yates and colleagues (Reference 20) developed a technique for proteomic analysis called Multidimensional Protein Identification Technique (MudPIT). In this technique the proteins in a proteome are first digested into relatively small peptides using protease enzymes including trypsin. These peptides are then fractionated

by ion-exchange chromatography by increasing the salt concentration in steps. In each step, the salt concentration is raised slightly and the least strongly bound peptides are eluted onto reversed-phase beads. An acetonitrile gradient is run, eluting the peptides by hydrophobicity (Reference 20). In the MudPIT approach the ion-exchange and reversed-phase beads are packed into a narrow capillary in sequence and the eluent from the reversed-phase section is directed into an electrospray mass spectrometer (Figure 49). This is repeated multiple times producing very significant separation of the peptides from the proteomic digest (Figure 50).

The advantage of two-dimensional chromatography of peptides from digested proteomes is that breaking proteins into peptides allows separation and identification of proteins that are missed by gel electrophoresis, such as



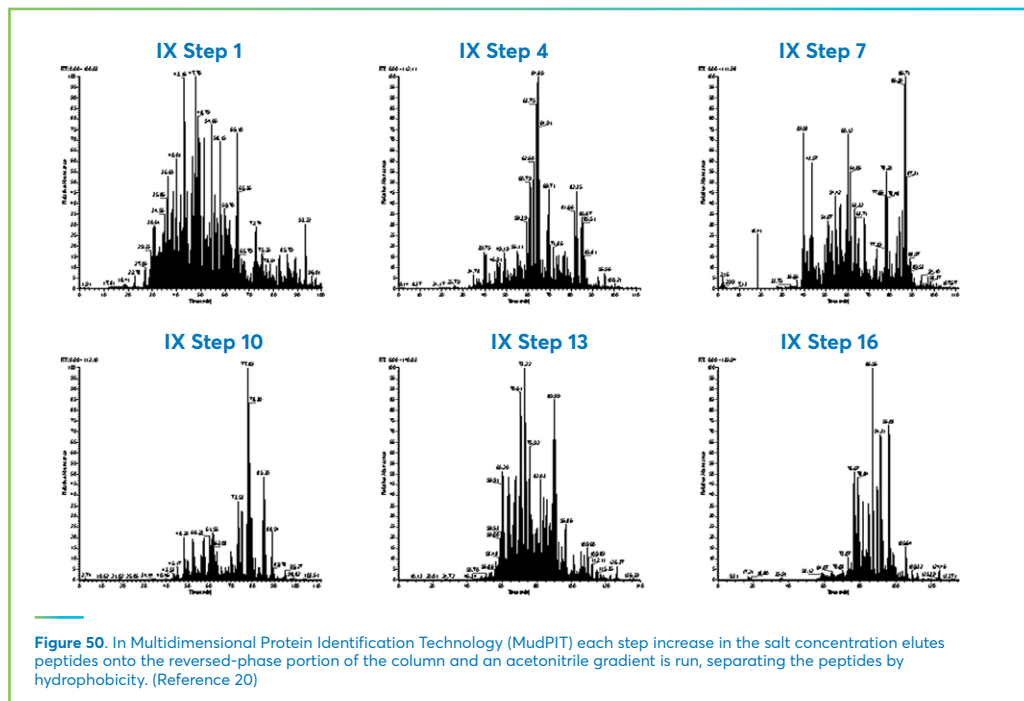
**Figure 49.** Multidimensional Protein Identification Technology (MudPIT) consists of an initial fractionation of a protease digestion of an entire proteome by ion exchange chromatography followed by a lengthy reversed-phase separation of the peptides from each fraction of the ion exchange separation. Narrow capillaries are packed with ion exchange beads and reversed-phase beads and the eluent from the chromatography is directed into an electrospray mass spectrometer.

hydrophobic, membrane-bound proteins and low abundance proteins. The disadvantage is that information on PTMs is usually lost in the MudPIT approach. Also, many more peptides are formed than there are proteins, on average 20 - 50 peptides per protein, which must be separated.

More recent advances have used affinity chromatography, immobilized metal affinity (IMAC) and lecithin affinity chromatography to further isolate sub-fractions of peptides either prior to ion-exchange separation or at later stages.

**Protein identification.** Proteins separated by gel electrophoresis are generally digested with proteases (trypsin in most cases) and the resulting peptides are analyzed by MALDI mass spectrometry and identified by reference to protein databases.

Peptides separated by MudPIT and related approaches are directed to an electrospray mass spectrometer and the peptides identified by MS and MS/MS with the use of protein databases. When a few peptides from a given protein are identified, the protein itself can be identified.





# References

## Introduction

01. Reversed-Phase High Performance Liquid Chromatography of Insulins from Different Species, J. Rivier and R. McClintock, *J. Chrom.* 268, 112-119 (1983)
02. Peptide Mapping of Subtilisins as a Practical Tool for Locating Protein Sequence Errors during Extensive Protein Engineering Projects, T. Christianson and C. Paech, *Anal. Biochem.* 223, 119-129 (1994)
03. Retention Model for Proteins in Reversed-Phase Liquid Chromatography, X. Geng and F.E. Regnier, *J. Chrom.* 296, 15-30 (1984)

## Columns

04. The Importance of Silica Type for Reverse-Phase Protein Separations, J.D. Pearson, N.T. Lin and F.E. Regnier, *Anal. Biochem.* 124, 217-230 (1982)
05. Analysis of peptide mixtures by capillary high performance liquid chromatography: A practical guide to small-scale separations, M.T. Davis and T.D. Lee, *Protein Science* 1, 935-944 (1992)
06. Microscale Structure Analysis of a High Molecular Weight Hydrophobic Membrane Glycoprotein Fraction with Platelet-Derived Growth Factor-Dependent Kinase Activity, P. Tempst, D. Woo, D.B. Teplow, R. Aebersold, L.E. Hood and S.B.H. Kent, *J. Chrom.* 359, 403-412 (1986)

## Analytical Conditions

07. Purification of Integral Plasma Membrane Proteins by Reverse-Phase High Performance Liquid Chromatography, M.R. Sussman, *Anal. Biochem.* 169, 395-399 (1988)
08. Subunit Analysis of Bovine Cytochrome c Oxidase by Reverse-Phase High Performance Liquid Chromatography, N.C. Robinson, M.D. Dale and L.H. Talbert, *Arch. of Biochem. and Biophys.* 281(2), 239-244 (1990)
09. Separation and Purification of S49 Mouse Lymphoma Histones by Reversed-Phase High Performance Liquid Chromatography, M.C. McCroskey, V.E. Groppi and J.D. Pearson, *Anal. Biochem.* 169, 427-432 (1987)
10. Computer simulation as a tool for the rapid optimization of the high-performance liquid chromatographic separation of a tryptic digest of human growth hormone, R.C. Chloupek, W.S. Hancock and L.R. Snyder, *J. Chrom.* 594, 65 - 73 (1992)
11. Increasing the Sensitivity of UV Detection in Protein and Peptide Separations when using TFA-Acetonitrile Gradients, G. Winkler, *LC/GC* 5(12), 1044-1045 (1987)
12. Temperature selectivity effects in reversed-phase liquid chromatography due to conformation differences between helical and non-helical peptides, Y. Chen, C.T. Mant and R.S. Hodges, *J. Chrom* 1010, 45-61 (2003)

13. Temperature as a variable in reversed-phase high-performance liquid chromatographic separations of peptide and protein samples. I. Optimizing the separation of a growth hormone tryptic digest, W.S. Hancock, R.C. Chloupek, J.J. Kirkland and L.R. Snyder, *J. Chrom.* 686, 31-43 (1994)
14. Development of an Analytical reversed-phase high-performance liquid chromatography - electrospray ionization mass spectrometry method for characterization of recombinant antibodies, T.M. Dillon, P.V. Bondarenko and M.S. Ricci, *J. Chrom.A* 1053, 299 - 305 (2004)
17. Enhanced sensitivity for peptide mapping with electrospray liquid chromatography-mass spectrometry in the presence of signal suppression due to trifluoroacetic acid-containing mobile phases, A. Apffel, S. Fischer, G. Goldberg, P.C. Goodley and F. Kuhlmann, *J. Chrom. A.* 712, 177 - 190 (1995)
18. Overview of Protein and Peptide Analysis by Mass Spectrometry, Section 16.1.14 in *Current Protocols in Protein Science*, published by John Wiley & Sons
19. High-flow ion spray liquid chromatography/mass spectrometry, G. Hopfgartner, T. Wachs, K. Bean, and J. Henion., *Anal. Chem.* 65, 439 - 446 (1993)

### HPLC - Mass Spectrometry

15. Evaluation of Genetic Stability of Recombinant Human Factor VIII by Peptide Mapping and On-line Mass Spectrometric Analysis, M.J. Besman and D. Shiba, *Pharm. Research* 14(8), 1092 - 1098 (1997)
16. Application of capillary electrophoresis, high-performance liquid chromatography, on-line electrospray mass spectrometry and matrix-assisted laser desorption ionization-time of flight mass spectrometry to the characterization of single-chain plasminogen activator, A. Apffel, J. Chakel, S. Udiavar, W. Hancock, C. Souders, E. Pungor, Jr., *J. Chrom.* A. 717, 41 - 60 (1995)

20. Direct Analysis of Protein Complexes Using Mass Spectrometry, A. Link, J. Eng, D. Schieltz, E. Carmack, G. Mize, B. Garvik, and J. Yates III., *Nature Biotech.* 17, 676-682 (1999)

### Applications






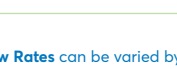
21. Sample Preparation for Peptide Mapping - A Pharmaceutical Quality Control Perspective, N. Lundell and T. Schreitmuller, *Anal. Biochem.* 266, 31- 47 (1999)
22. Rapid analytical tryptic mapping of a recombinant chimeric monoclonal antibody and method validation challenges, K. Kannan, M.G. Mulkerrin, M. Zhang, R. Gray, T. Steinharter, K. Sewerin, R. Baffi, R. Harris and C. Karunatililake, *J.Pharm. and Biomed. Analysis* 16, 631 - 640 (1997)

23. Stability of Peptides and Proteins, C. Schoneich, M.J. Hageman and R.T. Borchardt, in *Controlled Drug Delivery*, 205 - 228 (1997), publisher American Chemical Society
24. The Role of Quality Control in Biotechnology: An Analytical Perspective, R.L. Garnick, N.J. Solli and P.A. Papa, *Anal. Chem.* 60, 2546-2557 (1988)
- 25.25 Chemical Instability of Proteins, S. Li, C. Schoneich and R. Borchardt, *Pharmaceutical News* 2(5), 12 - 16 (1995)
26. Deamidation of asparagine and glutamine residues in proteins and peptides: structural determinants and analytical methodology, R. Bischoff and H.V.J. Kolbe, *J. Chrom. B.* 662, 261 - 278 (1994)
27. A Reversed-Phase High-Performance Liquid Chromatographic Method for Characterization of Biosynthetic Human Growth Hormone, R.M. Riggin, G.K. Dorulla and D.J. Miner, *Anal. Biochem.* 167, 199 - 209 (1987)
28. Selective Deamidation of Ribonuclease A, A.D. Donato, M.A. Ciardiello, M. de Nigris, R. Piccoli, L. Mazzarella and G. D'Alessio., *J. Biological Chem.* 268, 4745 - 4751 (1993)
29. Formation of Isoaspartate at Two Distinct Sites during in Vitro Aging of Human Growth Hormone, B.A. Johnson, J.M. Shirokawa, W.S. Hancock, M.W. Spellman, L.J. Basa and D.W. Aswad, *J. Biological Chem.* 264, 14262 - 14271 (1989)
30. Chemical Instability of Protein Pharmaceuticals: Mechanisms of Oxidation and Strategies for Stabilization, S. Li, C. Schoneich and R. Borchardt, *Biotechnology and Bioengineering* 48, 490 - 509 (1995)
31. Characterization of Methionine Sulfoxide Derivatives of Recombinant Coagulation Factor VIIa, T. Kornfelt and L. Palm, *WCBP II Conference* (1998), Poster P-1104T
32. Separation of Oxidized Human Growth Hormone Variants by Reversed-Phase High Performance Liquid Chromatography, G. Teshima and E. Canova-Davis, *J. Chrom.* 625, 207 - 215 (1992)
33. Model of Protein Conformation in the Reversed-Phase Separation of Interleukin-2 Muteins, M. Kunitani and D. Johnson, *J. Chrom.* 371, 313-333 (1986)
34. Oxidative Refolding of Insulin-like Growth Factor 1 Yields Two Products of Similar Thermodynamic Stability: A Bifurcating Protein-Folding Pathway, J. Miller, et.al., *Biochem.* 32, 5203-5213 (1993)
35. Strategies for Locating Disulfide Bonds in a Monoclonal Antibody via Mass Spectrometry, R. Mhatre, J. Woodard and C. Zeng., *Rapid Comm. in Mass Spec.* 13, 2503-2510 (1999)
36. Reversed-Phase Chromatography of Interleukin-2 Muteins, M. Kunitani, P.A. Hirtzer, D. Johnson, R. Halenback, A. Boosman and K. Koths, *J. Chrom.* 359, 391 - 402 (1986)

### Protein/Peptide Purification

37. Reversed-Phase High Performance Liquid Chromatography: Preparative Purification of Synthetic Peptides, J. Rivier, R. McClintock, R. Galyean, and H. Anderson, *J. Chrom.* 288, 303-328 (1984)
38. Practical Aspects of Preparative Reversed-Phase Chromatography of Synthetic Peptides, C.A. Hoeger, R. Galyean, R.A. McClintock and J.E. Rivier, in *HPLC of Peptides and Proteins: Separation, Analysis and Conformation*, Pages 753-764
39. Process Purification of Polypeptides and Proteins by Reversed-Phase Column Chromatography: Misconceptions and Reality, P. Lu, C.D. Carr, P. Chadwick, M. Li and K. Harrison., *BioPharm* 14 (9), 28 - 35 (2001)
40. Factors affecting the separation and loading capacity of proteins in preparative gradient elution high-performance liquid chromatography, Y-B. Yang, K. Harrison, D. Carr and G. Guiochon, *J. Chrom.* 590, 35-47 (1992)
41. A Simple and Rapid Purification of Commercial Trypsin and Chymotrypsin by Reversed-Phase High-Performance Liquid Chromatography, K. Titani, T. Sasagawa, K. Resing and K.A. Walsh, *Anal. Biochem.* 123, 408-412 (1982)
42. Erythropoietin purification, Por-Hsiung and T. Strickland, U.S. Patent No. 04667016
43. Expression, Purification and Characterization of Recombinant Murine Granulocyte-Macrophage Colony-Stimulating Factor and Bovine Interleukin-2 From Yeast, V. Price, D. Mochizuki, C.J. March, D. Cosman, M.C. Deeley, R. Klinke, W. Clevenger, S. Gillis, P. Baker and D. Urdal, *Gene* 55, 287-293 (1987)
44. Production of Recombinant Human Colony Stimulating Factors in Yeast, S. Gillis, D. Urdal, W. Clevenger, R. Klinke, H. Sassenfeld, V. Price and D. Cosman, *Behring Institute Research Communications* 83, 1-7 (1988)
45. Production Scale Purification of Biosynthetic Human Insulin By Reversed-Phase High-Performance Liquid Chromatography, E.P. Kroeff, R.A. Owens, E.L. Campbell, R.D. Johnson and H.I. Marks, *J. Chrom.* 461, 45-61 (1989)

## APPENDIX: COLUMN CHARACTERISTICS

	Column Diameter (mm)	Typical low Rate (ml/min)	Sample Capacity (µg)	Maximum Practical Load
<b>Capillary</b> 	0.075 0.15 0.30 0.50	0.00025 0.001 0.005 0.01	0.1 0.2 1 2	- - - -
<b>Microbore</b> 	1.0	0.05	10	-
<b>Narrow Bore</b> 	2.1	0.2	50	-
<b>Analytical</b> 	4.6	1.0	200	10 mg
<b>Semi-Preparative</b> 	10	5.0	1 mg	50 mg
<b>Preparative</b> 	21.2	10-30	5 mg	200 mg

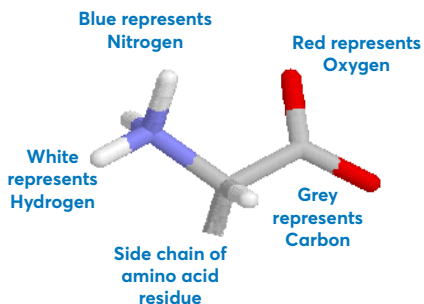
**Flow Rates** can be varied by a factor of 2 without loss of resolution

**Sample Capacity** is the amount of polypeptide that can be chromatographed without loss of resolution

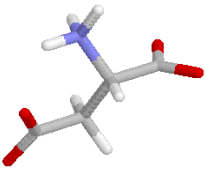
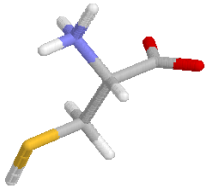
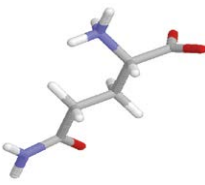
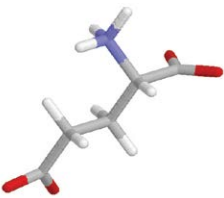
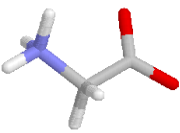
**Maximum Practical Sample Load** is the maximum amount of polypeptide that can be purified with reasonable yield and purity

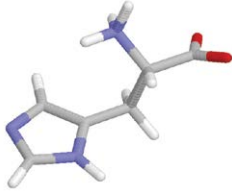
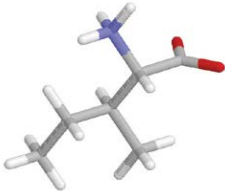
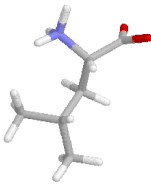
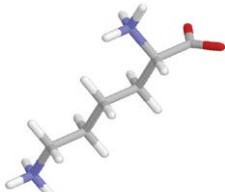
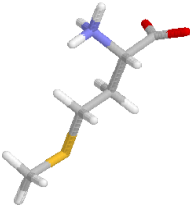
## APPENDIX: AMINO ACID CHARACTERISTICS

Amino acids are amphoteric, meaning that they contain both an acidic group (carboxylic acid) and a basic group (primary amine). These, however, are condensed with leading and following residues in a peptide bond. A series of amino acids connected via peptide bonds form a polypeptide backbone. The characteristics of a protein/peptide are determined by the side chains attached to the amino acid core. These are described and their functions outlined in this appendix. The name, three letter abbreviation and single letter code are included.

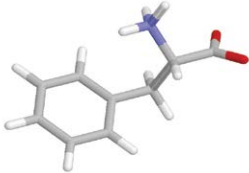
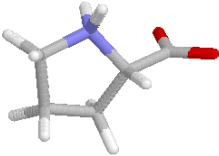
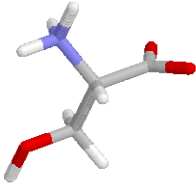
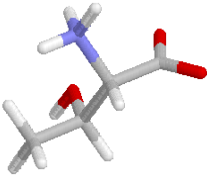
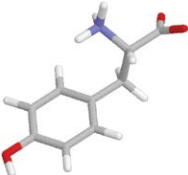


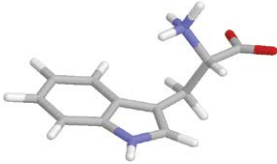
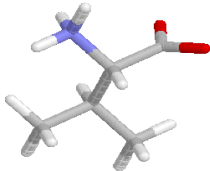
Amino Acid	Characteristics	Structure
<b>Alanine</b> Ala <b>A</b>	<b>Type:</b> Hydrophobic <b>Function:</b> The side chain consists of a methyl group, which is hydrophobic. As with other hydrophobic amino acids, alanine assists with forming and maintaining tertiary structure.	
<b>Arginine</b> Arg <b>R</b>	<b>Type:</b> Basic <b>Function:</b> The side chain has three nitrogens closely grouped, making Arg the most basic amino acid with a pK of ~12.5. Arginine forms electrostatic complexes with aspartic or glutamic acid (salt bridges), strengthening tertiary structure. It also interacts with acidic groups being bound to a protein, such as the phosphate groups of nucleic acids.	
<b>Asparagine</b> Asn <b>N</b>	<b>Type:</b> Slightly polar <b>Function:</b> Asparagine is distinguished as being the primary location for deamidation and for being the point of attachment of glycans (oligosaccharides).	

Amino Acid	Characteristics	Structure
<b>Aspartic Acid</b> Asp D	Type: Acidic Function: Aspartic acid has a pK of ~4 and interacts with basic amino acids forming electrostatic complexes (salt bridges). Asp is also the binding site for positively charged groups such as metal ions, required by many proteins to be functional.	
<b>Cysteine</b> Cys C	Type: Slightly polar (sulphur-containing) Function: Cysteine plays a major role in protein tertiary structure by reacting with other cysteine side chains and forming a covalent linkage (disulphide bond) between different parts of a polypeptide chain.	
<b>Glutamine</b> Gln Q	Type: Slightly polar Function: Glutamine does not deamidate as easily as asparagine and therefore is rarely a site of deamidation.	
<b>Glutamic Acid</b> Glu E	Type: Acidic Function: Glutamic acid has a pK slightly greater than 4 and interacts with basic amino acids forming electrostatic complexes (salt bridges). Glu is also the binding site for positively charged groups such as metal ions required by many proteins to be functional.	
<b>Glycine</b> Gly G	Type: Neutral Function: Glycine is very small with a side chain of only a hydrogen. It is a neutral extender within the backbone.	

Amino Acid	Characteristics	Structure
<p><b>Histidine</b> His H</p>	<p>Type: Basic Function: The side chain is slightly basic with a pK of 6-7. Histidine can bind acids but binding depends on the environmental pH.</p>	
<p><b>Isoleucine</b> Ile I</p>	<p>Type: Hydrophobic Function: As with other hydrophobic amino acids, isoleucine residues assist with forming and maintaining tertiary structure.</p>	
<p><b>Leucine</b> Leu L</p>	<p>Type: Hydrophobic Function: As with other hydrophobic amino acids, leucine assists with forming and maintaining tertiary structure.</p>	
<p><b>Lysine</b> Lys K</p>	<p>Type: Basic Function: The side chain is a primary amine and has a pK of ~10.5. Lysine forms electrostatic complexes with aspartic or glutamic acid (salt bridges), strengthening tertiary structure. It also interacts with acidic groups being bound to a protein, such as the phosphate groups of nucleic acids.</p>	
<p><b>Methionine</b> Met M</p>	<p>Type: Slightly hydrophobic (sulphur-containing) Function: Methionine is the most likely amino acid to be oxidized. It acts as an indicator of oxidation. Methionine oxidation may also affect biological activity.</p>	



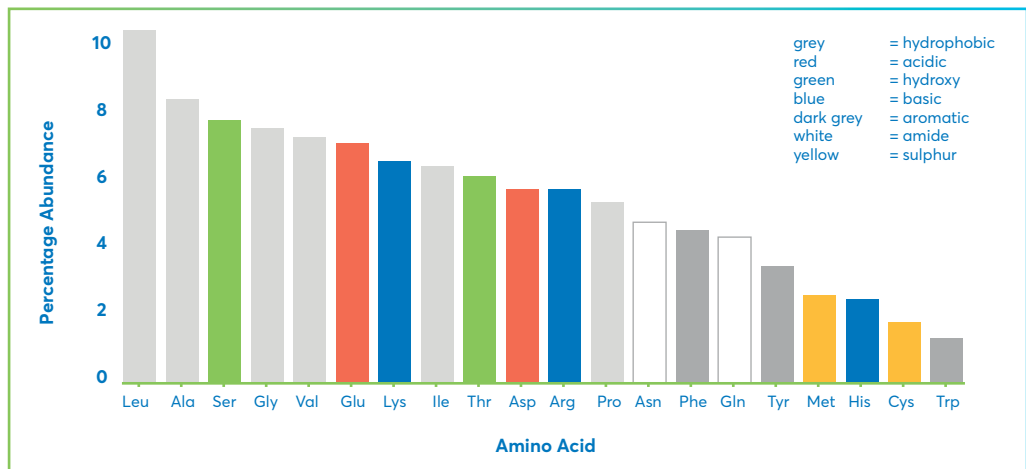
Amino Acid	Characteristics	Structure
<p><b>Phenylalanine</b></p> <p><b>Phe</b></p> <p><b>F</b></p>	<p>Type: Hydrophobic and aromatic.</p> <p>Function: As with other hydrophobic amino acids, phenylalanine assists with forming and maintaining tertiary structure. Being aromatic, it is also a useful detection target by UV light absorption.</p>	
<p><b>Amino Acid</b></p> <p><b>Proline</b></p> <p><b>Pro</b></p> <p><b>P</b></p>	<p>Type: Slightly hydrophobic</p> <p>Function: Proline is unique in that its hydrophobic side chain is connected to the amino terminus. This creates a strain in the polypeptide backbone, tending to bend the backbone. This is a structural characteristic assisting in the formation of tertiary structure.</p>	
<p><b>Serine</b></p> <p><b>Ser</b></p> <p><b>S</b></p>	<p>Type: Polar (hydroxy)</p> <p>Function: Serine is able to hydrogen bond with other residues or parts of the protein. It also can hydrogen bond with other proteins or bound groups. Serine is also a site of phosphorylation on cellular proteins.</p>	
<p><b>Threonine</b></p> <p><b>Thr</b></p> <p><b>T</b></p>	<p>Type: Polar (hydroxy)</p> <p>Function: Threonine is able to hydrogen bond with other residues or parts of the protein. It also can hydrogen bond with other proteins or bound groups. Threonine is also a site of phosphorylation on cellular proteins.</p>	
<p><b>Tyrosine</b></p> <p><b>Tyr</b></p> <p><b>Y</b></p>	<p>Type: Polar (hydroxy) and aromatic</p> <p>Function: Tyrosine is able to hydrogen bond with other residues or parts of the protein. It also can hydrogen bond with other proteins or bound groups. Tyrosine is also a site of phosphorylation on cellular proteins. Tyrosine is aromatic and provides a UV absorption site.</p>	

Amino Acid	Characteristics	Structure
<b>Tryptophan</b> Trp W	Type: Hydrophobic and aromatic Function: Tryptophan has a multi-ring system and can hydrogen bond to a small extent. It is aromatic and provides a site of UV light absorption. Tryptophan also has a slight natural fluorescence.	
<b>Valine</b> Val V	Type: Hydrophobic Function: The side chain of valine is a hydrophobic isopropyl group. As with other hydrophobic amino acids, valine assists in forming and maintaining tertiary structure.	

**Abundance of individual amino acids.**

Hydrophobic amino acids are critical in forming tertiary structure and are often found at the core of a protein. They represent approximately 50% of the amino acids in a protein. Polar amino acids

have many important structural and functional purposes. They are quite abundant. The least abundant amino acids are methionine, histidine, cysteine and tryptophan.





# Setting science in motion to create a better world

---

Avantor® is a leading global provider of mission critical products and services to customers in the biopharma, healthcare, education & government, and advanced technologies & applied materials industries. We operate in more than 30 countries and deliver an extensive portfolio of products and services.