Drug Development and Industrial Pharmacy, 23(9), 915-927 (1997)

Analysis of Peptides and Proteins: Evaluation of Purity, Stability, and Structural Characterization of Insulin

A. Oliva, J. Fariña, and M. Llabrés

Departamento de Ingeniería Química y Tecnología Farmacéutica, Facultad de Farmacia, Universidad de La Laguna, 38200 La Laguna, Tenerife, Spain

ABSTRACT

The analysis of peptides and proteins is complicated either by their inherent reactivity (as physical change is produced in peptide structure due to aggregation) or by adsorption of sample material onto container surfaces. The characterization of proteins requires the use of several techniques; each probe measures a particular structural or functional feature. Methods for evaluating the purity and stability of peptides and proteins exist as well as various methods for their structural characterization: size, shape, and conformational changes. These are modifications of secondary and tertiary structures produced in the molecule during preparation, manipulation, and storage.

INTRODUCTION

In the last decades with the advances in recombinant DNA technology and chemical synthesis, the commercial production of proteins for pharmaceutical purposes has become feasible. However, with the rapid growth of these products, a demand for their analysis has been increasing. The identity and purity of these products must be assayed before the products are administered to patients. The development and production of peptides as drugs has presented a scientific challenge. Moreover, the analytical methods used to characterize and evaluate this group of substances will not only help our understanding of the substances' chemical and biological properties, but also enhance regulatory approval and safety of the resulting product.

The analysis of peptides is complicated by their inherent reactivity. The most common reaction is hydrolysis of amide linkages (mainly the side chain amide of a Gln or Asn residue (1)), although oxidation reaction, racemization, and polymerization are also frequently observed (2).

On the other hand, physical changes in peptide structure due to aggregation or adsorption of sample material onto container surfaces can result in an apparent loss of analyte as well as in yielding a new chemical entity.

Physicochemical identity and purity assessment are sufficient to guarantee therapeutic potency for small molecules; however, due to their complexity, the proteins require an additional evaluation of their activity. Purity must reflect not only chemical composition, but also chirality, conformation, and physical states of the aggregation of substances. Sometimes, a loss of purity may not affect biological activity, whereas in other cases, limitations on the scope of available analytical methodology cause impurities. These impurities, which are not recognized, can affect biological activity. The characterization of small molecules is a relatively singular work, while the characterization of proteins requires the use of several techniques, each probe measuring a particular structural or functioning feature. First, it is necessary to evaluate the stability and purity of this type of product using analytical methods such as chromatography and electrophoresis. Then several techniques are used for the determination of peptide amino acid composition and their sequences, (i.e., the primary structure). Other methods are used to determine the size, shape, and conformation, such as circular dichroism, mass spectrometry, resonance magnetic nuclear imaging, x-ray diffraction, and light scattering.

It is our intention to provide a review of current analytical methodology, especially in relation to peptides and proteins. In particular, insulin will be discussed because of its therapeutic interest and importance in pharmaceutical industry.

PURITY AND STABILITY

The determination of the purity of insulin requires the use of several analytical methods such as gel filtration and disk electrophoresis. Other methods include higher sensibility techniques such as radioimmunoassays (RIA), high-performance liquid chromatography (HPLC), and more recently, high-performance capillary electrophoresis (HPCE).

ELECTROPHORESIS

Electrophoresis can be defined as the controlled movement of charged species through a solvent in the presence of an electron field. It has been an effective method for the separation and characterization of macromolecules in function of their net charge or molecular size. Commonly used supports are paper, cellulose acetate, and gels such as polycrylamide and, in the last five years, fused-silica capillaries of different internal diameter. With the appropriate analytical conditions, this method can be used for evaluation of purity and stability of peptides and proteins.

Since the introduction of disk electrophoresis in polyacrylamide gels by Ornstein and Davis (3,4), this method has been widely used for characterization of pure insulin. Although various electrophoretic systems are used by different authors, they are all modifications of the original system by Ornstein and Davis. The more frequent conditions were the following: 10–200 μg of protein was applied at a slightly alkaline pH, using 20% acrylamide gel; and Coomassie Brilliant Blue (a staining system) was used. However, variations in acrylamide concentration, load per tube, content of dissociating agent (urea), and dye (for staining) have been described (5–8).

This method can be used to estimate the monodesamido insulin content in a semiquantitative manner by comparing gels containing known amounts of impurities added to monocomponent insulin. Comparisons are also made with a series of gels to which varying amounts of insulin have been applied as a protein standard (9). However, variations in the staining and poor band resolution have made this method less accurate and reproducible.

Isoelectric focusing separates components based on their isoelectric points (pIs) in the presence of a pH gradient. It is most often performed on a thin tube of polyacrylamide. The pH gradient, normally in the pH range of 3 to 10, is established with a mixture of low molecular weight synthetic polyaminopolycarboxilic acids (ampholytes). The optimal ampholyte mixture should have component pI closely over the desired pH range for peptide analysis.

However, it should be noted that both electrophoretic methods may not always be applicable to the analysis of peptide formulations, due to interference in peptide separation by excipients (10). Other limitations are their unsuitability for automation and the difficulty of removing separated components from supports for further characterization as well as the fact that staining methods frequently used in these systems do not offer adequate sensitivity. An alternative detection system with enhanced sensitivity may require radiochemical, fluorometric, or immunochemical markers.

High-performance capillary electrophoresis (HPCE) represents the first electrophoretic technique where the separation of analytes is performed inside a capillary tube. HPCE uses fused-slica capillary with an internal diameter of 25–100 μm and length of 50–100 cm. Ultraviolet (UV) absorbance is the universal method of

detection, although mass spectrometry and fluorescence detection are also used on a limited basis (11). The narrow diameter columns provide capillary electrophoresis with several advantages such as short analysis times, high separation efficiency, and minimal sample volume requirements.

The basic isoelectric focusing capillary electrophoresis (CE-IEF) is similar to isoelectric focusing in gel, but it is performed in a coated capillary, usually with polyacrylamides that are chemically bonded to capillary walls. The CE-IEF system is simple and eliminates all of the difficulties of gel electrophoresis. The principal advantages of this method are its automation, fast analysis (< 30 min), high resolution (0.1 pIs), linearity over a wide pH gradient (pH 3–10), and direct quantification. Proteins and peptides with a wide range of pIs and a broad range of molecular weights can now be analyzed in a single run.

Micellar electrokinetic capillary chromatography (MECC), which uses an ionic micellar solution as the separation solution, is another mode of capillary electrophoresis. Capillary electrophoresis (CE) is a separation technique of ionic analytes only, whereas MECC is capable of separating both ionic and non-ionic analytes. Most advantages of CE apply to MECC as well, and many applications of MECC separations have been reported (12). The MECC is based on the differential partitioning of an analyte between the micelle (which is a pseudo-stationary phase) and the aqueous phase; therefore, the choice of surfactants and modifiers of the aqueous phase is important for manipulation separation selectivity (13). The major advantages of MECC compared with those of CE are that positively charged, negatively charged, and neutral species can all be separated in the same run.

On the other hand, in isotachophoresis, another type of capillary electrophoresis, the sample is introduced in a capillary tube between an electrolyte of higher mobility (leading electrolyte) and one of lower mobility (terminating electrolyte). The leading, terminating, and sample ions all have net charges of the same sign. After applying an electric field, sample components of different mobility will move through the capillary tube, where those of highest mobility travel the greatest distance in a given period. The sample zones can be detected with various methods, including thermometry, electrical conductivity, and ultraviolet detection. For example, a basic peptide such as secretin is typically characterized in a protonated form by cationic isotachophoresis at pH 5.1, with potassiunm cations (k^+) as the leading electrolite and β-alanine as the terminating electrolyte (10).

Today, capillary electrophoresis surpasses every aspect of conventional gel electrophoresis, but especially in simplicity, speed, high resolution, automation, selectivity, reproducibility, and short analysis times. Thus, the capillary allows on-line detection for precise and accurate quantification of proteins as well as peptides. Separations of macromolecules obtained by capillary electrophoresis offer a unique and different selectivity with respect to high-performance liquid chromatography or gas chromatography; therefore, it is an ideal supplementing of these techniques. Moreover, the resolution, selectivity, and analysis times can be improved by modifying the voltage, buffer, ionic strength, and pH, by adding micelles, and by chemical modifications of the capillary wall.

The application of a combined system of HPLC and HPCE (CE and MECC) for the step-by-step control of recombinant human insulin production has been suggested by Klyushnichenko et al. (14). In the near future, this combination of different methods can form the basis of a production control system.

CHROMATOGRAPHY

Chromatography compasses a wide variety of separation methods based on the relative affinities of a solute for the mobile and stationary phase. Several chromatographic methods have been used in the characterization of pharmaceutical peptides, including thin-layer chromatography (TLC) and gas chromatography (GC). TLC has been used for monitoring impurities in synthetic peptide preparations such as those found in salmon calcitonin (15) and to characterize other peptides of therapeutic interest such as glucagon, angiotensins, and insulin (16). GC has mainly been used for the identification of small molecule impurities, and cannot be used to identity peptides containing more than a few peptides because of their nonvolatility and thermal lability. However, GC has an important application in peptide sequence determination when performed in conjunction with mass spectrometry.

In the past decades, both techniques have become displaced by high-performance liquid chromatography (HPLC). Today, HPLC is probably the most versatile and widely applied method for separation and identification of peptides, replacing or at least supplementing more laborious methods. The main advantages of HPLC are speed, simplicity, short analysis times, and automation. Table 1 shows the mains applications of HPLC in the analysis of peptides and proteins.

Table 1

Major Applications of HPLC Used in the Control of Purity and Stability of Peptides and Proteins

Amino acid analysis
Purity control of synthetic peptide
Identification of degradation products, especially macro-aggregate forms of proteins
Molecular weight determination
Fingerprint analysis
Sequence determination
Isolation, purification, and control of products obtained by genetic engineering
Isolation of protein fragments from sequencing studies

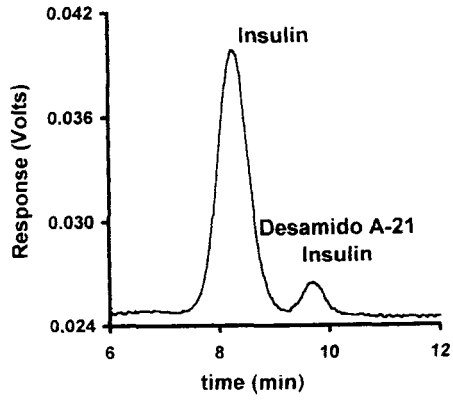

Figure 1. Reverse-phase HPLC separation of human insulin and their respective A-21 desamido insulin (27).

The most recent method for the analysis of insulin is HPLC in reverse phase mode (RP-HPLC), with a non-polar matrix as stationary phase, and a buffer mixture and organic solvent as mobile phase. RP-HPLC packing usually consists of hydrocarbon chain chemically bonded to microparticulate silica gels; chain lengths frequently employed are the octyl (C-8) and octadecyl (C-18) forms. This packing is compatible with gradient elution systems and permits a greater flexibility in the selection of mobile phase composition. The more commonly used buffers are mono- or dihydrogenphosphates, sulphates, and monohydrogencarbonates at pH <3, and the more frequently used organic solvent is acetonitrile in a percentage between 20–30% of total composition.

The separation of bovine and porcine insulin was the first important application of HPLC in insulin chemistry (17–19) followed by the separation of human insulin from insulin from two other species (20). HPLC methods for separating several species including chicken, rabbit, sheep, and horse have been published (21,22). A number of HPLC methods for the quantification of monodeamido insulin have been published (19,23–27). In most cases, the composition and monodesamido content of the species can be determined in the same analysis. Many authors (25–27) have described HPLC methods as being capable of separating bovine, porcine, and human insulin, and their respective monodesamido derivatives in the same analysis (Fig. 1). However, few publications have described the use of HPLC for quantification of insulin and its derivatives in insulin preparations. Oliva et al. (27) have proposed an RP-HPLC method for quantification of insulin in stability studies of commercial human insulin preparations during storage at different conditions.

RP-HPLC is also used for fingerprint analysis of the proteolytic enzyme digestion of protein, since one way to characterize a macromolecule is to describe the fragments that result from a given reaction. The protein is digested by the action of a protease into peptide fragments and the size and nature of these fragments are characteristic of the parent protein. By separating and counting the fragments, a well-known pattern emerges—a "fingerprint"—in proteins. The analytical method most often used for separation of these fragments is the RP-HPLC technique with gradient elution and UV detection. In fact, with this technique it is possible to verify the identity of a protein or peptide, allowing for insulin to be easily distinguished. (25,28–31). Table 2 shows the characteristic of frequent protease used to check the identity of insulin. For insulin, the *Staphylococcus aureus* V8 protease appears ideal since it gives rise to the formation of a reasonable number of well-sized fragments. In this case, native insulin should be cleaved into only 4 fragments, since it disposes of a rapid separation and detection method, and the identification of individual fragments is possible, i.e., by mass spectrometry. Figure 2 shows RP-HPLC pattern of *Staphylococcus aureus* fingerprint of human insulin. The fingerprint method allows the analysis of bovine, porcine, and human insulin (28–31), the identification of modified insulin (7,29), and the characterization of different degradation products of the insulin during storage under different conditions (28,32).

Due to the increased use of HPLC in insulin analysis during the past years, little doubt remains that HPLC will be one of the fundamental methods for analysis and characterization of insulin purity and stability in the future.

Size-exclusion HPLC (SE-HPLC) or gel filtration is one type of chromatography where the separation is

Table 2

Characteristics of Some Proteases Used to Verify the Identity of Proteins

Protease	Specificity
Trypsin	C-terminal of Arg or Lys residues
α-Chymotrypsin	C-terminal of Phe, Trp, or Tyr residues
Pepsin	C-terminal of Phe, Trp, or Tyr residues
S. aureus	C-terminal of Glu residues
Thermolysine	N-terminal of Leu, Ile, Val, Met, or Phe residues
Papain	Relatively nonspecific

based on molecular size. Molecules will be eluted from the column in order to decrease molecular weight.

Steiner (33,34) showed by means of gel filtration that commercial insulin preparations, purified solely by crystallization, contained impurities with higher molecular weight than insulin. These impurities were later identified to be mainly proinsulin, proinsulin intermediates, and covalent insulin dimer. Since, several gel filtration systems for the analysis of insulin have been described (35,36). Common to all methods are the use of Bio-Gel P30 (Bio-Rad Laboratories) or Sephadex G-50 (Pharmacia Fine Chemicals) and acetic acid as eluent.

The original material used for SE-HPLC of proteins was coated controlled-porosity glasses, but adsorption was a serious problem. SE-HPLC became operational in 1980 with the development of uniform rigid particles which possessed relatively uniform pores that were sufficiently large to be permeated by polypeptide molecules. Silica-based supports with different hydrophilic coatings were introduced for SE-HPLC of proteins and peptides for the first time by Welinder in 1980 (37). Currently, there is a high chromatographic supports trade in the market to aid in the analysis for SE-HPLC.

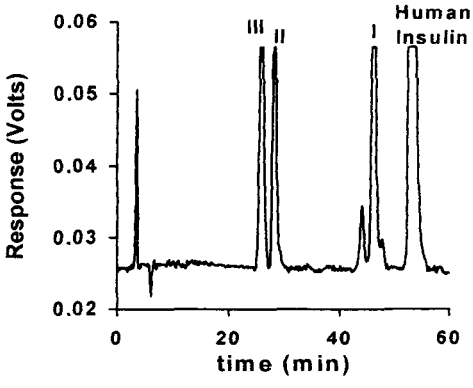

Figure 2. Reverse-phase HPLC pattern of *Staphylococcus aureus* fingerprint of human insulin (28).

However, one important aspect in the choice of SE-HPLC column is that there should be no interaction between the peptide and the surface of the stationary phase. The problem is especially common with silica-based materials but can be minimized by bonding the surface with modifiers; therefore, the choice of mobile phase is extremely important. Columns containing support materials having a pore size of 5 to 400 nm are readily available and several different bonded silica materials are routinely used. Other column supports are TSK-PW, which is a hydroxilated polyether copolymer, and Superose, which is a cross-linked agarose derivative. These two supports are not silica based, but are stable over a much wider pH range. The polymeric supports allow very efficient protein separations, particularly in the high molecular weight range.

Although size exclusion is the principal factor governing retention in the column, there are several types of interaction, mainly of ionic, electrostatic, hydrophobic, and mixed modes as well as adsorption to the gel (which also affects elution behavior) (38). For example, the size of the protein is influenced by the ionic strength of the solvent (39). Therefore, the elution time with size exclusion columns depends also on the salt concentration in the mobile phase. The amount of salt not only influences the conformation of the protein (and for this reason the hydrodynamic volume), but it can also influence the chromatographic separation mechanism itself on columns (40).

In practice, it is very difficult to achieve a purely size-exclusion effect in the separation of components, although several modes can be adopted to reduce this effect. For example, hydrophobic interaction can be usually reduced by the addition to the eluent of denaturating agents, such as sodium dodecyl sulphate (SDS), urea, guanidine hydrochloride, acetonitrile, or amino acids (38,41–43).

SE-HPLC is often used for qualitative and quantitative analysis of proteins and peptides in both wide and

narrow ranges of molecular weight, and can be performed in both native and denaturating conditions (44–46). SE-HPLC has been used for the analysis of high-molecular impurities of insulin, which are formed due to polymerization and aggregate formation during purification, manipulation, and storage (47).

Welinder (37) performed the analysis of insulin and insulin-related compounds by SE-HPLC, using three different silica-based chromatographic supports. Currently, Klyushnichenko and Wulfson (38) and Brange et al. (48) have proposed several SE-HPLC methods for the separation of insulin, proinsulin, covalent insulin dimer, and macroaggregates. Oliva et al. (49) have also proposed a SE-HPLC system for the separation of insulin and covalent insulin dimer in pharmaceutical preparations using a silica-based supports column (Fig. 3).

STRUCTURAL CHARACTERIZATION

There are several methods for the identification of amino acid sequences (i.e., the primary structure), but all methods involve the partial or total digestion of the peptide sample. One of these methods is Edman degradation. Other methods are for studying the size, shape, and conformational changes (i.e., modifications of secondary and tertiary structure) such as circular dichroism, mass spectrometry, resonance magnetic nuclear imaging, x-ray diffraction, and light scattering.

DETERMINATION OF PRIMARY STRUCTURE

A complete amino acid sequence is often required to unequivocally establish the chemical identity of the protein, which involves a total cleavage of all peptide linkages in the molecule and the separation and quantitation of the released amino acid. A complete hydrolysis of protein can be performed in acid or basic conditions or enzymatic digests. Acid hydrolysis consists of dissolving the peptide in 6N HCl and heating at 110°C for 24 hr. This type of hydrolysis leads to a total destruction of tryptophan residues. However, it is often possible to estimate tryptophan content through the use of base hydrolysis. Usually, this hydrolysis involves heating a peptide solution in 2N NaOH at 100°C for several hours. This procedure does not harm tryptophan, but does result in a complete or partial degradation of various amino acids such as cystine, serine, treonine, and arginine and produces amino acid racemization. Finally, proteolytic enzymes can also be used to promote peptide hydrolysis. These enzymes cleave peptide bonds on the amino acid side of residues located at strategic positions within the sequence.

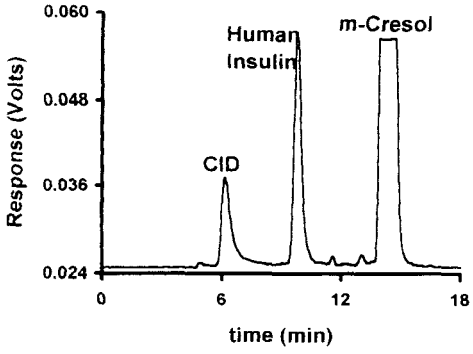

Figure 3. Size-exclusion HPLC separation of human insulin and insulin-related compounds corresponding to commercial human insulin preparation stored at 60°C for 25 days. CID: covalent insulin dimer (49).

Generally, the separation and identification of amino acid can be achieved by thin layer chromatography, gas chromatography, and especially by reversed-phase HPLC with precolumn derivatization and ultraviolet detection. For higher sensitivity, fluorescence instead of UV detection is essential. The formation of *o*-phthalaldehyde, dansyl, dabsyl, phenylthiocarbamyl, phenylthiohydantoin (Edman degradation), and butylthiocarbamyl derivatives of amino acids has been widely used in RP-HPLC (50). Derivatization of amino acids with phenylisothiocyanate (PITC), which has been used in the Edelman degradation, is an excellent precolumn derivatization method especially for secondary amino acids, proline, and hydroxyproline (50). Excess from PITC and those of its substitutes during derivatization of PTC amino acids must be completely removed. This was done using a high vacuum system in order to avoid interferences during the analysis, although extraction procedures (with diethylether, heptane, cyclohexane, or hexane as extractant) were recommended instead of the high vacuum conditions (51).

Quantitation of the amino acid is performed chromatographically. Gas chromatography was perhaps the more widely used method for this purpose. However, in the recent years, the reversed-phase HPLC has been the current method of choice. Usually, the separation is performed on an octadecylsilane C-18 column using a gradient of increasing concentration of an organic solvent such as methanol or acetonitrile. Molnàr-Perl (51) has made a review of HPLC conditions suggested for the analysis of PTC amino acids.

For many proteins of biological interest having between 200 to 400 residues, it is not always possible to determine the amino acid sequence directly. In these cases, the polypeptide chain must be cleaved into smaller fragments either enzymatically or chemically, and chromatographically separated, purified, and sequenced. Table 3 shows the partial cleavages of proteins using enzymatic or chemical methods.

Another way to characterize a protein is to describe the fragments which result from an enzymatic hydrolysis. By separating and counting the fragments, a well-known pattern emerges as a "fingerprint" in proteins. The analytical method most often used for separation of these fragments is RP-HPLC technique with gradient elution and UV detection. This technique must not only permit the identification of a protein or peptide, but also distinguish, i.e., insulin, among different species (25,28–30).

TECHNIQUES USED FOR THE DETERMINATION OF SECONDARY AND TERTIARY STRUCTURES

Circular Dichroism

Circular dichroism (CD) is the difference in the absorbance of two opposite forms of polarized light. If circularly polarized left and right light beams that are equal in amplitude and phases are combined, the result will be plane polarized light. If these two light beams are passed through a substance that is optically active, one will be preferentially absorbed. The combination of the two unequal beams will result not in plane polarized light, but in light that is elliptically polarized. The ellipticity of the polarized light can be quantified by the ratio of the major and minor axes of the ellipse; this is what is measured in CD.

CD is a form of optical spectroscopy that is extremely sensitive to protein structure. The technique can be used to study the secondary and tertiary structures of proteins and nucleic acid, and to give information about structural changes that occur during molecular interaction or by factors such as temperature, pH, dissociating agent, etc.

CD is generally measured between 180 and 250 nm, which is commonly known as the far-UV region. The CD spectrum of this region depends on the secondary structure conformation of the peptide groups. The rest of the CD of proteins originate mainly from aromatic side chains and disulfide groups—tyrosine, tryptophan, hystidine, and phenylalanine—all of which absorb between 240 and 300 nm and so can contribute to the CD spectrum in this region known as the near-UV. Thus, CD measured in the far-UV region reflects the organization of the polypeptide backbone (i.e., secondary structure), while the spectrum in the near-UV region is more closely related to the environment around the aromatic side chains and their relative orientation (tertiary structure) (52).

The CD spectrum of native insulin shows negative minima at 208, 222, and 275 nm (Fig. 4). Every negative minima reflects an aspect of insulin structure. Therefore, it can display the following three bands:

208-nm band: This band may be attributed to α-helix (53) which is a characteristic feature of mono-

Table 3

Partial Cleavages of Proteins Performed Enzymatically or by Chemical Methods

Enzymatic Cleavages	Specificity
Trypsin	C-terminal side of Arg or Lys
α-Chymotrypsin	C-terminal side of Phe, Trp, or Tyr
Pepsin	C-terminal side of Phe, Trp, or Tyr
S. aureus	C-terminal side of Glu
Thermolysine	N-terminal side of Leu, Ile, Val, Met, or Phe
Papain	Relatively nonspecific
Chemical Cleavages	
Cyanogen bromide	C-terminal side of Met
o-Iodosobenzoic acid	C-terminal side of Trp
Hydroxylamine	Amido linkages of Asp-Gly
2-Nitro-5-thiocyanobenzoic acid	N-terminal side of Cys

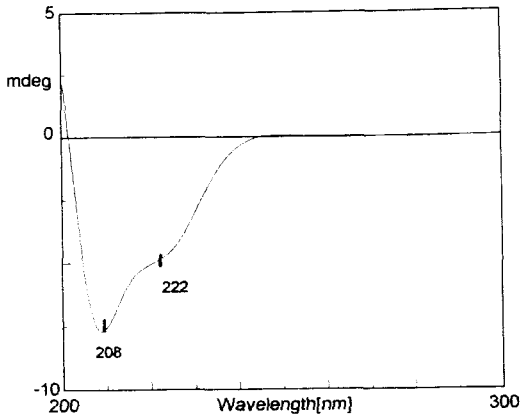

Figure 4. Far-UV circular dichroism spectrum of human insulin in 0.05 N HCl (Oliva et al., unpublished results).

mer insulin. It is not affected by sample dilution or modified insulin which may aggregate in a manner different from native insulin. However, the band is attenuated by the addition of zinc ions or other divalent metal ions and may be increased in the presence of some additives.

222-nm band: This band can be assigned to β-structure (54), which is a predominant feature of the dimer. The 222-nm band is attenuated by dilution of sample and in those derivatives that either do not aggregate or do so in a manner different from insulin (55).

275-nm band: This band was assigned to tyrosine and phenylalanine aromatic residues (55–58). Attenuation of this band has been associated with disaggregation of insulin (55), while strengthening of the band is associated with conditions that enhance insulin aggregations. The optical activity of the aromatic residues contributing to this band is therefore dependent upon the state of insulin self-association (57).

Insulin exhibits a complex association behavior in crystal and in solution. In the crystalline state, it exists as a hexamer and in solution, it exists as an equilibrium between monomers, dimers, tetarmers, hexamers, and possibly some higher associated states (57–59). It is possible to study these states by CD. The CD spectrum of insulin depends on many features such as the concentration, pH, zinc ion, or other divalent ion concentrations, presence of certain additives, and so on. For example, the ellipticity at 275 nm decrease with the decreasing concentration, but when the concentration is below 1.0 μm, the measurement is difficult due to the noise problem. In the far-UV region, the ellipticity at 208 and 222 nm decreases with the concentration, but the change at 208 nm is less than that of 222 nm. It is interesting to note that, while the ellipticities at 208 and 222 nm both decrease with the concentration, the ratio of ellipticities increases with decreasing concentration (59–61). This ratio was used by authors to calculate the percentage of monomer insulin in solution in function of the concentration.

The CD spectrum of insulin changes with pH. The ellipticity at 222 nm changes at pH acid, while it remains mostly unchanged at pH basic (59). This change in the CD spectrum could be due to the dissociation of dimers into monomers. A similar feature occurred at 275 nm since the presence of certain divalent metal ions (Fe^{+2}, Cu^{+2}, Co^{+2}) causes an increase of this band, and the aggregation of insulin is favored (55). However, Brems et al. (62) observed a decrease in ellipticity at 275 nm as a result of 20% ethanol addition due to the dissociation of insulin. Phenolic additives widely used for the preservation of insulin preparations can have a profound effect on the hormone's conformation in solution. An addition of moderate amounts of *m*-cresol, one of the most frequently used preservatives in insulin preparation, results in marked changes of insulin's CD spectrum. In the far-UV ellipticity, it is increased by 10–20% while the shape of the spectrum remains unaltered, whereas in the near-UV, the main band centred at 275 nm does not change (63). Therefore, the CD can serve as a probe to detect small, often functionally important conformational changes in the protein structure. The CD spectrum of insulin can give information about the insulin self-association in solution; however, the CD spectrum depends on many factors and its possible application in pharmaceutical preparations is limited due to the presence of phenolic additives because these produce important alterations in the CD spectrum.

X-ray Diffraction

This is one of the most powerful techniques available for the study of protein structure. X-ray diffraction is basically a scattering phenomenon. If crystals of the pure protein are available, the three dimensional conformation of the molecule can be described unequivocally. Moreover, it is possible to get a resolution as low as 1.5 to 0.5 Å in some cases.

The structural information obtained by x-ray diffraction can be used as reference for other structural methods such as circular dichroism. For example, the attenu-

ation of the near-UV CD of insulin samples by the addition of 20% ethanol observed by Brem et al. (62) is consistent with the x-ray crystallography data which show that hydrophobic bonding is the main source of interaction between the multimers (64).

Extensive x-ray crystallographic studies of insulin were conducted and have provided detailed information concerning its secondary, tertiary, and cuaternary structures. Today, the x-ray diffraction is a support and confirmation method of structural changes (mainly in modified insulins) detected for other methods such as circular dichroism, nuclear magnetic resonance imaging, and light scattering (64–71).

Nuclear Magnetic Resonance (NMR)

In recent decades, development of NMR spectrometers coupled with the use of Fourier transform processes in data adquisition has extended the capability of NMR to analyze constituents of macromolecules, including amino acids within peptides. A significant advantage of NMR is that it represents a nondestructive method of amino acid identification as well as improving the quantitation of amino acids that might be partially destroyed by hydrolytic procedures or that cannot easily be derivatized for sensitive detection. An important disadvantage of proton NMR with respect to characterization of macromolecules is that relatively small ranges of ^{1}H chemical shifts (0 to 10 ppm) coupled with the large peak widths (5 to 30 Hz) result in broad bands of overlapping resonances. However, a higher resolution can be achieved using two dimensional J resolved spectroscopy since this probe enhances NMR spectral dispersion by spreading spin-controlled multiplets into a second dimension and suppressing the line broadening due to inhomogeneity (10).

Human insulin could be studied by high resolution NMR, although it can vary in function of the insulin self-association and with analysis conditions. Kline and Justice (72) obtained the ^{1}H-NMR for human insulin using low pH and the addition of acetonitrile to overcome the protein's self-association, while Roy et al. (73) worked under conditions of concentration, ionic strength, and pH basic, under which human insulin is monomeric. Moreover, these authors demonstrated that the native insulin monomer has a conformation that differs in some respects from the conformation of higher aggregates. Weiss et al. (74) performed the ^{1}H-NMR studies of insulin as a function of acetic acid concentration. In 20% acetic acid, the proteins were found to be monomeric at the milimolar concentration required for NMR study. The dimerization occurs with increasing protein concentration or temperature changes in 20% acetic acid. These factors produced modifications of ^{1}H-NMR spectra, which are consistent with those expected from the UV and circular dichroism results.

^{13}C-NMR is potentially more valuable than ^{1}H-NMR for studying individual amino acids in a peptide and giving more information about the molecular structure than the chemical environments. Hruby et al. (75) have described the use of this technique using vasopressin, oxytocin, and a series of related analogues. However, the low natural abundance of the ^{13}C isotope involves using samples with concentrations even higher than those required for ^{1}H-NMR and longer data adquisition time to obtain useful information. For these reasons, the use of this technique for structural characterization of peptides and proteins is very limited.

Mass Spectrometry (MS)

In the last 20 years, mass spectrometry (MS) has been used to determine the molecular weight and structural characterization of biopolymers such as proteins, peptides, polysaccharides, and polynucleotides. MS is a powerful analytical technique. Compared with other traditional methods of polymer analysis, today's mass spectrometer can directly measure the molecular mass of oligomer and the substructures of macromolecules (76). In addition, new ionization techniques allow macromolecules to be efficiently ionized and vaporized without degradation, constituting some of the reasons for the increasingly important analysis of biopolymers by MS. Two of the more widely used and newest ionization techniques are the matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI). MALDI uses photons to make gaseous ions directly from solid material, and ESI produces gaseous ion sprays directly from solutions.

Currently, many types of mass analyzers are available for MS, although the major differences between these mass analyzers are the power of resolution and the upper mass range. Analyzers include the magnetic sector, quadrupole ion trap, time of flight (TOF), and Fourier transform (FT) ion cyclotron resonance mass analyzers.

Electrospray Ionization

In the late 80's, analysts began to use electrospray exclusively, and it has revolutionized the field of MS in many ways. In this case, a sample solution is sprayed

or nebulized under the influence of a high electric field. The resulting aerosol is desolvated by a combination of heat, gas flow, and vacuum. With electrospray, ions are typically formed as multiple charged species. This multicharging produces a near Gaussian distribution of peaks corresponding to the different mass/charge ratios. With this type of source, the more widely used analyzers are the quadrupole and sector magnetic obtaining a mass accuracy of 0.01% or better, subpicomole detection limits, and a mass range that is greater than 100,000 Daltons (77). Although Fourier transform ion cyclotron resonance mass spectrometry can also be used to analyze large macromolecules or complex mixtures, quadrupole mass analyzers are generally used with ESI sources because they are readily available, inexpensive, and offer sufficient mass range and resolution. Moreover, ESI has the advantages that it can be coupled to liquid chromatographic techniques including size-exclusion chromatography and capillary electrophoresis (76,77).

Meng et al. (78) established the molecular mass of several proteins such as insulin, cytochrom C (horse), lysozyme, myoglobin, trypsin inhibitor, α-chymotrypsin, and alcohol deshydrogenase. Until 1990, researchers had reported the analysis of more than 100 peptides and proteins by ESI mass spectrometry.

MALDI

Matrix-assisted laser desorption ionization (MALDI) relies on the use of a solid chomophoric matrix, chosen to absorb laser light, which is co-mixed with the analyte (79). Biopolymers are mixed with an excess of matrix in an aqueous solution and deposited on a probe. The solvent is allowed to evaporate before inserting the probe into the mass spectrometer. When the pulsed laser beam strikes, the sample surfaces are desorbed/ionized at high efficiency. Thus, the simple addition of laser absorbing matrices has led to the detection of singly and doubly protonated biopolymers with masses greater than 300,000 Daltons (77). The resulting mass spectra usually contain protonated or deprotonated molecules of the analyte with virtually no fragment ions.

MALDI requires the use of matrices that absorb laser radiation at the wavelength used for desorption, and has been used for the analysis of proteins with masses greater than 200,000 Da. A variety of wavelengths are suitable for desorption, and ultraviolet, visible, and infrared radiation have been employed for this purpose with several types of matrices (80). Table 4 shows the more frequently used matrices in MALDI mass spectrometry. The time of flight (TOF) and Fourier transform (FT) mass analyzers for laser desorption are commonly used. However, FT mass spectrometer are more sophisticated since they provide high resolution, accurate mass measurement, and multistage MS capabilities. MALDI/TOF provides cost-effective measurements for low molecular masses (less than 1000 Da) to over 250,000 Da. The laser desorption FTMS technique can provide accurate molecular mass information for polymer that is less than 10,000 Da and provide chemical information on the molecular structure. However, the TOF mass analyzer is the more widely used since it has some advantages such as ease of preparation, low sample constitution, and short analysis times (76,77,80). Table 5 shows several peptides and proteins analyzed by MALDI mass spectrometry.

Table 4

More Frequently Used Matrices in MALDI Mass Spectrometry

Spectro Region	Matrix
Ultraviolet	Nicotinic acid
	2-5-Hydroxybenzoic acid
	Cinnamic acid and derivatives
	3-Nitrophenol
Visible	Mixtures of nitrobenzyl alcohol and Rodamina 6G
Infrared	Nicotinic acid
	Glycerol
	Succinic acid
	Urea

Perhaps one of the more studied peptides has been insulin. Many authors (80–82) have determined its molecular mass, separating the bovine and porcine insulin in the same analysis. On the other hand, Solouki and Russell (83) were able to determine the molecular weight of insulin dimer.

Therefore, mass spectrometry is a powerful analytical tool in the increasingly important analysis of biopolymers. It is capable of providing detailed qualitative and quantitative information compared to other analytical methods. In the near future, higher resolution techniques will enable mass spectrometry to accurately detect large biopolymers to within 1 Da. Moreover, the possibility of coupling to liquid chromatography system and capillary electrophoresis will become an invaluable

Table 5

Peptides and Proteins Analyzed by MALDI Mass Spectrometry

	Molecular Mass	Analyser	Reference
Gramicidín S	1142	FT[a]	Hettich & Buchanan, 1991
Gramicidín D	1882	FT	Hettich & Buchanan, 1991
Bovine/insulin A chain	2535	FT	Castoro & Wilkins, 1993
Bovine insulin	5734	FT	Köster et al., 1992
Porcine insulin	5778	FT	Hunt et al., 1987
β-Lactoglobulin A	18277	TOF[b]	Karas & Hillenkamp, 1988
Porcine trypsin	23463	TOF	Karas & Hillenkamp, 1988
Bovine albumin	67000	TOF	Karas & Hillenkamp, 1988

[a]Fourier transform
[b]Time of flight

tool for the quality control of recombinant protein (76,77,84).

Light Scattering

Light scattering (LS) is one of the most commonly used techniques for determining the molecular weight of synthetic polymers, although polydispersity and its size can also be determined. However, in the recent years, LS has become popular for its applications in the biopolymers field, i.e., in the characterization of proteins.

Currently, analysts can determine several parameters such as molecular weight, particle size, radius of gyration, degree of aggregation, and conformational change produced by several factors. When a solution of a macromolecule is illuminated with laser light, the amount of light scattered by the macromolecule can be measured. In static or Rayleigh light scattering, the intensity of the scattered light is measured as a function of the angle of observation and solute concentration. The result gives direct information on molecular weight and dimensions of the macromolecule. Thus, in dynamic light scattering, the fluctuations of the scattered radiation are used to determine the diffusion coefficient of the molecule. This information may be used to calculate the size of the molecule using the Stokes-Einstein equation. The fluctuations of scattering density are directly related to the random motion of the molecules in solution. Large molecules diffusing slowly produce slow fluctuations, while small molecules that move faster cause more rapid fluctuations in the scattered light. These fluctuations can be followed and analyzed by means of the autocorrelation function (85).

Today, all light-scattering detectors can be interfaced with other types of detectors such as UV-VIS (in the case of proteins) or refractive index detector (in the case of polymers) as well as with the liquid chromatography system (83,86,87). Many applications of interfacing HPLC with LS have been published; however, the newest application has been centerd on the characterization of proteins and their aggregates. The LS detection does not alter the nature of the species during detection, and the percentage and nature of aggregates truly represent the sample content in solution (88).

On the other hand, Rayleigh LS can also provide information about the shape and size of a molecule in terms of the radius of gyration (Rg), whose value is important in studying the denaturation of proteins and protein interactions since larger Rg values involve denaturation proteins (88).

Bohidar and Geissler (89) have studied the self-association behavior in zinc–insulin solution as a function of time at neutral pH and at room temperature using both static and dynamic light scattering. Baudys et al. (90) used the dynamic light scattering to determine the apparent diffusion coefficient and the hydrodynamic diameter of native insulin and insulin derivatives under different conditions. These authors established that the hydrodynamic diameter is enhanced when the insulin self-association states are higher (5.6 nm for insulin hexamer and 2.6 nm for insulin monomer).

ACKNOWLEDGMENT

This research has been financed by Comisión Interministerial de Ciencia y Tecnologia (CICYT) as part of projects SAF 94-1379.

REFERENCES

1. K. Patel and R. T. Borchardt, Pharm. Res., 7, 703 (1990).
2. A. M. Oliva, J. B. Fariña, and M. Llabrés, Industria Farmacéutica, 1, 33 (1995).
3. L. Ornstein, Disc electrophoresis-I. Ann. N.Y. Acad. Sci. 121, 321 (1964).
4. B. J. Davis, Disc electrophoresis-II. Ann. N.Y. Acad. Sci. 121, 404 (1964).
5. J. Schlichtkrull, J. Brange, A. H. Christiansen, O. Hallund, L. G. Heding, K. H. Jorgensen, S. M. Rasmusen, E. Sorensen, and A. Volund, Metab. Res. (Suppl. Ser.) 5, 134 (1974).
6. A. H. Reiner, P. Nemes, and C. Bucholtz, Anal. Biochem., 64, 509 (1975).
7. R. E. Chance, E. P. Kroeff, J. A. Hoffmann, and B. H. Frank, Diabetes Care, 4, 147 (1981).
8. J. Brange, L. Langkjaer, S. Havelund, and A. Volund, Pharm. Res., 9, 715 (1992).
9. *British Pharmacopoeia*, 1980.
10. C. S. Randall, T. R. Malefyt, and L. A. Sternson, in *Peptide and Protein Drug Delivery: Approaches to the Analysis of Peptides*, Marcel Dekker, New York, 1991, p. 203.
11. K. D. Altria and M. T. Kersey, LC-GC International., 8, 201 (1995).
12. S. Terabe, in *Capillary Electrophoresis Tecnology* (N. Guzman, ed.), Marcel Dekker, New York, 1993, p. 65.
13. S. Terabe, N. Matsubara, Y. Ishihama, and Y. Okada, J. Chromatogr., 608, 23 (1992).
14. V. E. Klyushnichenko, D. M. Koulich, S. A. Yakimov, K. V. Maltsev, G. A. Grishina, I. V. Nazimov, and A. N. Wulfson, J. Chromatogr., 661, 83 (1994).
15. P. H. Corran and D. H. Calam, in *Recent Developments in Chromatography and Electrophoresis* (Frigerio and Renoz, eds.), Elsevier, Amsterdam, 1979, p. 341.
16. L. Lepri, P. G. Desideri, and D. Heimler, J. Chromatogr., 211, 29 (1981).
17. M. E. Biemond, W. A. Sipman, and J. Olivié, J. Liq. Chromatogr., 2, 1407 (1979).
18. U. Damgaard and J. Markussen, Horm. Metab. Res., 11, 580 (1979).
19. A. Dinner and L. Lorenz, Anal. Chem., 51, 1872 (1979).
20. S. Terabe, R. Konaka, and K. Inouye, J. Chromatogr., 172, 163 (1979).
21. M. Ohta, H. Tokunaga, T. Kimura, H. Satoh, and J. Kawamura, Chem. Pharm. Bull. (Tokyo), 31, 3566 (1983).
22. J. Rivier and R. McClintock, J. Chromatogr., 268, 112 (1983).
23. G. Szepesi and M. Gazdag, J. Chromatogr., 218, 597 (1981).
24. L. F. Lloyd and P. H. Corran, J. Chromatogr., 240, 445 (1982).
25. N. A. Farid, L. M. Atkins, G. W. Becker, A. Dinner, R. E. Heiney, D. J. Miner, and R. M. Riggin, J. Pharm. Biomed. Anal. 7, 185 (1989).
26. R. Gregory, S. Edwards, A. Nigel, and M. S. Yateman, Diabetes Care, 14, 42 (1991).
27. A. Oliva, J. Fariña, and M. Llabrés, Indus. Farmac., 2, 43 (1996).
28. A. Oliva, J. Fariña, and M. Llabrés, Drug Dev. Ind. Pharm., 23, 127 (1997).
29. U. Grau, Diabetes, 34, 1174 (1985).
30. M. Vestling, J. Chem. Edu., 68, 958 (1991).
31. R. J. Schilling and A. K. Mitra, Pharm. Res., 8, 721 (1991).
32. R. Darrington and B. Anderson, J. Pharm. Sci., 84, 275 (1995).
33. D. F. Steiner and P. E. Oyer, Proc. Natl. Acad. Sci. USA, 57, 473 (1967).
34. D. F. Steiner, O. Hallund, A. Rubenstein, S. Cho, and C. Bayliss, Diabetes, 17, 725 (1968).
35. R. L. Rolando and D. Torroba, Experientia, 28, 1169 (1972).
36. B. V. Fisher and P. B. Porter, J. Pharm. Pharmacol., 33, 203 (1981).
37. B. S. Welinder, J. Liq. Chromatogr., 3, 1399 (1980).
38. V. E. Klyuschnichenko and A. N. Wulfson, Pure and Appl. Chem., 65, 2265 (1993).
39. M. Herold, Int. Lab., 3, 34 (1993).
40. M. Potschka, J. Chromatogr., 441, 236 (1988).
41. J. E. Rivier, J. Chromatogr.,, 202, 211 (1980).
42. G. D. Swergold and C. S. Rubin, Anal. Biochem., 131, 295 (1983).
43. G. J. Murray, R. J. Youle, S. E. Gandy, G. C. Zirzow, and J. A. Barranger, Anal. Biochem., 147, 301 (1985).
44. Y. B. Yang and M. Verzele, J. Chromatogr., 391, 383 (1987).
45. Y. Kato, K. Komiya, H. Sasaki, and T. Hashimoto, J. Chromatogr., 193, 29 (1980).
46. T. Imamura, K. Konishi, M. Yokoyama, and K. Konishi, J. Liq. Chromatogr., 4, 613 (1981).
47. E. Schrader and E. F. Preiffer, J. Liq. Chromatogr., 6, 1121 (1985).
48. J. Brange, S. Havelund, and P. Hougaard, Pharm. Res., 9, 727 (1992).
49. A. Oliva, J. Fariña, and M. Llabrés, M. Int. J. Pharm., 143, 163 (1996).
50. K. L. Woo and S. H. Lee, J. Chromatogr., 667, 105 (1994).

51. I. Molnár-Perl, J. Chromatogr., 661, 43 (1994).
52. P. G. Varley, Microbiol. Eur., 3, 37 (1993).
53. M. J. Ettinger and S. N. Timasheff, Biochemistry, 10, 824 (1971).
54. F. Quadrifoglio and D. W. Urry, J. Am. Chem. Soc., 90, 2760 (1968).
55. J. Goldman and F. H. Carpenter, Biochemistry, 13, 4566 (1974).
56. J. W. Morris, D. A. Mercola, and E. R. Arquilla, Biochim. Biophys. Acta., 160, 145 (1968).
57. S. Sato, C. D. Ebert, and W. K. Sun, J. Pharm. Sci., 72, 228 (1983).
58. J. Brange, U. Ribel, J. Hansen, G. Dodson, M. Hansen, S. Havelund, S. Melberg, F. Norris, K. Norris, L. Snel, A. Sorensen, and H. Voigt, Nature 333, 679 (1988).
59. Y. Pocker and S. B. Biswas, Biochemistry, 19, 5043 (1980).
60. S. P. Wood, T. L. Blundell, A. Wollmer, N. R. Lazarus, and R. W. Neville, Eur. J. Biochem., 55, 531 (1975).
61. G. Schwartz, G. T. Burke, J. D. Chanley, and P. G. Katsoyannis, Biochemistry, 22, 4561 (1983).
62. D. N. Brems, P. L. Brow, L. D. Heckenlaible, and B. H. Frank, Biochemistry, 29, 9289 (1990).
63. A. Wollmer, B. Rannefeld, B. R. Johansen, K. R. Hejnaes, P. Balschmidt, and F. B. Hansen, Biol. Chem., 368, 903 (1987).
64. E. N. Baker, T. L. Blundell, J. R. Cutfield, E. J. Dodson, C. G. Dodson, D. M. Hodgking, and R. E. Hubbard, Phil. Trans. R. Soc., 319, 369 (1988).
65. M. J. Adams, T. L. Blundell, E. J. Dodson, G. G. Dodson, M. Vijayan, E. N. Baker, M. M. Harding, D. C. Hodking, B. Rimmer, and S. Sheat, Nature, 224, 491 (1969).
66. T. L. Blundell, S. M. Cutfield, E. J. Dodson, G. G. Dodson, D. C. Hodking, D. A. Mercola, and M. Vijayan, Nature, 231, 506 (1971).
67. G. Bentley, E. Dodson, G. Dodson, D. Hodking, and D. Mercola, Nature, 261, 166 (1976).
68. D. Wang, W. Chang, and J. Dai, Sci. Sin., B25, 603 (1982).
69. W. Chang, D. Stuart, and J. Dai, Sci. Sin., B29, 1273 (1986).
70. U. Derewenda, Z. Derrewenda, E. Dodson, G. Dodson, C. Reynolds, G. Smith, C. Sparks, and D. Swenson, Nature, 338, 594 (1989).
71. D. C. Hodgking and D. A. Mercola, in *Handbook of Physiology* (S. R. Geiger, ed.), Washington, D.C., 1972, p. 139.
72. A. D. Kline and R. M. Justice, Jr., Biochemistry, 29, 2906 (1990).
73. M. Roy, R. W. Lee, J. Brange, and M. F. Dunn, J. Biol. Chem., 265, 5448 (1990).
74. M. A. Weiss, B. H. Frank, I. Khait, A. Pekar, R. Heiney, S. E. Shoelson, and L. J. Neuringer, Biochemistry, 29, 8389 (1990).
75. V. J. Hruby, K. K. Deb, A. F. Spatola, D. A. Upson, and D. Yanomoto, J. Am. Chem. Soc., 101, 202 (1979).
76. J. C. Campana, L. S. Sheng, S. L. Shen, and B. E. Winger, Anal. Chem., 3, 239 (1994).
77. L. S. Krull, R. Mhatre, and J. Cunniff, LC-GC Int., 8, 14 (1995).
78. C. K. Meng, M. Mann, and J. B. Fenn, Paper presented at the 36th Annual Conference on Mass Spectrometry and Allied Topics, San Francisco, California, USA, June 5–10, 1988.
79. R. C. Beavis, Org. Mass. Spectrom., 27, 653 (1992).
80. J. A. Castoro and C. L. Wilkins, Anal. Chem., 65, 2621 (1993).
81. D. F. Hunt, J. Shabanowitz, J. R. Yates, N. Z. Zhu, D. H. Russell, and M. E. Castro, Proc. Natl. Acad. Sci. USA, 84, 620 (1987).
82. R. L. Hettich and M. V. Buchanan, J. Am. Soc. Spectrom., 2, 22 (1991).
83. T. Solouki and D. H. Russell, Proc. Natl. Acad. Sci. USA, 89, 5701 (1992).
84. D. Arnott, J. Shabanowitz, and D. F. Hunt, Clin. Chem., 39, 2005 (1993).
85. R. Pearlman and T. Nguyen, in *Peptide and Protein Drug Delivery* (V. H. Lee, ed.), Marcel Dekker, New York, 1991, p. 247.
86. H. Miklautz, J. Riemann, and H. J. Vidic, J. Liq. Chromatogr., 9, 2073 (1986).
87. T. Nicolai, J. Chromatogr., 389, 286 (1987).
88. I. S. Krull, R. Mathre, and J. Cunniff, LC-GC Int., 8, 83 (1995).
89. H. Bohidar and E. Geissler, Biopolymers, 23, 2407 (1984).
90. M. Baudys, T. Uchio, D. Mix, D. Wilson, and W. K. Sung, J. Pharm. Sci., 84, 28 (1995).
91. M. Karas and F. Hillenkamp, Anal. Chem., 60, 2299 (1988).