Electrospray Ionization Mass Spectrometric Analysis of Proteins

S. K. Chowdhury, V. Katta, and B. T. Chait

The Rockefeller University
New York, N.Y. 10021

INTRODUCTION

Electrospray is a 'soft' ionization technique in which intact gas phase ions of involatile and thermally labile biomolecules are produced directly from an analyte solution of interest at atmospheric pressure. Electrospray occurs when a strong electric field is applied to a small flow of liquid emerging through a fine capillary tube. The strong electric field causes the surface of the emerging liquid to become highly charged resulting in the formation of a fine spray of charged droplets. The solvents continuously evaporate from the droplets as the droplets proceed from atmospheric pressure to the vacuum leading to the formation of gas phase solute ions.

A unique feature of the ionization process is the production of multiply charged ions of biopolymers such as proteins with high efficiency. These multiply charged ions result from the attachment of protons and/or cations (e.g. Na^+) to the basic sites on the molecule. Because of multiple charging, the mass-to-charge ratios (m/z) of high molecular mass biopolymer ions can be small. For proteins the mass-to-charge ratios typically range between 500-2500. Therefore, conventional mass spectrometers with limited m/z range can be used to analyze large proteins. The subject of electrospray ionization (ESI) mass spectrometry of peptides and proteins has been reviewed [1,2].

Recently, we have constructed a simple, inexpensive, and easy-to-use atmospheric pressure electrospray ionization interface that we have coupled to a commercial quadrupole mass spectrometer. In the last year we have investigated a large number of peptides and proteins [3] and deduced structural information by mass spectrometric fragmentation [4] and by the peptide mapping technique [5]. The present article describes, in part, the materials presented at the NATO advanced research workshop held at the Minaki, Canada.

THE INSTRUMENT

A schematic representation of the Rockefeller University electrospray ionization mass spectrometer is shown in Figure 1. The instrument is comprised of two major components:

Methods and Mechanisms for Producing Ions from Large Molecules
Edited by K.G. Standing and W. Ens, Plenum Press, New York, 1991

a) the ion source and atmospheric pressure - vacuum interface and b) the quadrupole mass analyzer. The former component was constructed at the authors' laboratory, while the latter was purchased from Vestec Co., Houston, Texas. Since our initial description of the instrument [3], several small instrumental modifications and changes in experimental conditions were made. These will be described here.

The sample solution is electrosprayed from a hypodermic syringe needle, **a**, into the ambient air at a rate of 0.2 to 1.0 µl/min (Figure 1). The resulting highly charged droplets and solvated ions were transported into the vacuum of the mass spectrometer through a 20 cm long, 0.5 mm i.d. stainless steel capillary tube, **b**, as a result of the gas flow from the atmospheric pressure to the vacuum. The gas flow focuses the droplets and solvated ions towards the central axis of the tube and transports them efficiently into the vacuum chamber [6,7].

The capillary tube, **b**, passes through a concentric stainless steel metal sleeve wound with the heating tape, **e**. A thermocouple for monitoring the capillary tube temperature was fitted onto the sleeve. The sleeve around the capillary tube was heated to 80-150°C to assist in the evaporation of solvent molecules from the droplets that pass through it. Because the ions that exit the capillary tube are often strongly solvated, an electric field is applied between the capillary tube and the skimmer to remove solvent molecules by collisional activation (Fig. 1). The electric field in the region between the skimmer and the capillary tube can be easily controlled by varying the applied voltage on the capillary tube (the voltage on the skimmer is kept constant). At lower voltages desolvation of ions can be achieved [3], while at higher voltages it proves feasible to induce fragmentation of anlayte ions [4]. The combined (cumulative) effect of heat and collisional activation provides the total desolvation. The tube is normally heated to a fixed temperature between 80-150°C and the voltage on the capillary tube is varied to obtain the highest mass spectrometric response for a given ion. Such an optimization is performed by scanning the mass analyzer about a narrow m/z region and monitoring the ion signal of interest on the computer screen.

The desolvated ions/fragment ions that enter the mass analyzer chamber through the skimmer are m/z analyzed by the quadrupole mass analyzer and are detected by an off axis electron multiplier. The voltage applied to the front face of the multiplier is generally -3000 V for positive ions. No additional post acceleration was used.

The desolvation of ions/droplets in the present system is convenient and controllable, and does not require the use of the counter current flows of gas that have been employed by other workers [1,2]. The use of a metal capillary tube for transportation of ions and droplets in the present system reduces charging problems that are sometimes experienced in glass capillary tubes.

EXPERIMENTAL CONDITIONS

Preparation of Sample Solutions

The protein samples were weighed and were transferred into Eppendorf tubes. Appropriate amounts of distilled water were added to the samples and were sonicated for 5-10 minutes in an ultra-sonic bath to dissolve the sample completely. Highly purified methanol and acetic acid were then added slowly to the water solution to obtain a final sample concentration of 1-20 µM and solvent proportions of 40-50% water, 45-55% methanol, and 3-5% acetic acid. The solution was further sonicated for 5-10 minutes for complete dissolution of the analyte and was then electrosprayed for mass spectrometric analysis.

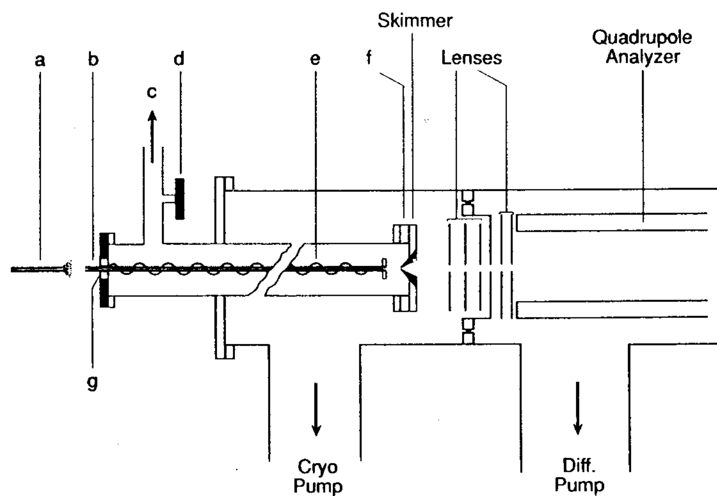

Figure 1 A schematic diagram of the electrospray ionization mass spectrometer (not drawn to scale). **a**, 150 um ID syringe needle at ~3.5 kV, **b**, 0.5mm ID, 200 mm long stainless steel capillary tube at 100-300 V, **c**, to rotary pump; **d**, pirani gauge; **e**, heating tape; **f**, Teflon insulating plate; **g**, Swagelock Teflon fitting. The skimmer (0.5 mm ID), lens elements, and quadrupole analyzer are from a Vestec 201 thermospray mass spectrometer.

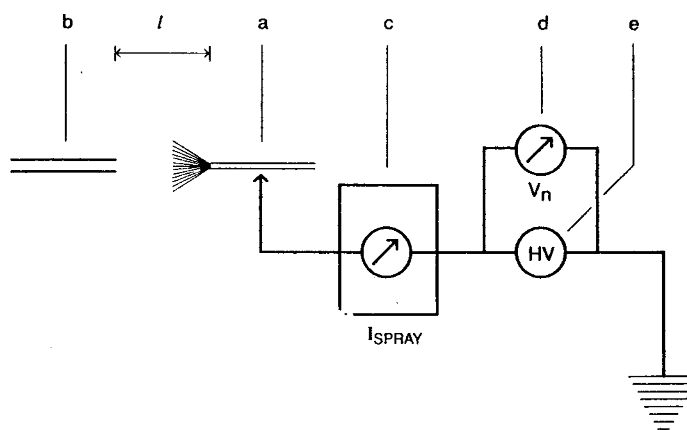

Figure 2 A schematic diagram of electrical setup for the electrospray unit. **a**, the spray needle; **b**, the metal capillary tube; **c**, an electrically isolated ameter to measure I_{spray}, the spray current; **d**, a voltmeter; and **e**, high voltage power supply unit for the needle.

About 20-50 μl of the analyte solutions were loaded into a 100 μl glass syringe (Hamilton, Reno, N.Y.) that has a 26 gage, stainless steel , 90° point needle (i.d. = 150 μm). A constant flow of liquid (0.2-2 μl/min) through the needle was maintained by a syringe pump (Harvard Apparatus, South Natick, MA; model 2400-001). The voltage applied to the needle, V_n, to obtain a stable electrospray is determined by the distance, *l*, between the needle tip, **a**, and the capillary tube, **b** (Figure 2) and the solvent composition of the analyte solution. For l = 0.4-1.0 cm and for the typical solvent compositions described above, a value of V_n = 3.5-4.5 kV is generally required to obtain a stable spray.

The spray conditions as a function of V_n are interesting. For l =0.55 cm and V_n = 3.3 kV a conical spray flume with a large solid angle is obtained (Figure 3). The central component of this flume contains large droplets. The spray current measured as shown in Figure 2 is small and is typically ~100 nA. The peptide/protein ion intensities determined by mass spectrometric analysis are also very small. When V_n is increased to ~3.8 kV a fine jet of spray with smaller solid angle (cone) is observed and the measured spray current increases to ~200 nA. Under this spray condition, stable peptide/protein ion signals are obtained with maximum sensitivity. When V_n is increased further to >4 kV, the spray splits into multiple diverse jets (Fig. 3) providing a spray current of >300 nA. However, the mass spectrometric sensitivity of ions decreases substantially and sometimes no trace of analyte ions could be found. A further increase in V_n causes an electrical breakdown producing a corona discharge between the spray needle and the transport capillary tube. The capillary tube also serves as a counter electrode for the spray needle. For peptide and protein analysis we use conditions that provide a fine, stable spray with maximum sensitivity as outlined above.

A particular advantage of the present electrospray system is that the electrospray is carried out in ambient air outside the vacuum housing and therefore the spray can be optimized by visually monitoring the spray conditions in addition to monitoring the spray current.

MASS SPECTRA AND SENSITIVITY

The typical electrospray ionization (ESI) mass spectra of peptides and proteins consist of a series of multiply charged ions, resulting from the attachment of protons to the available basic sites of protein/peptide molecules [1-3]. Figure 4 shows the ESI mass spectra of equine skeletal muscle myoglobin (Sigma cat. #M0630) between m/z 100 and 1400. The spectrum is a single scan acquired in 13 s from a 5 μM solution of myoglobin dissolved in water + methanol + acetic acid (45 : 52 : 3% v/v). The spectrum exhibits multiply charged ions arising from the attachment of 14 to 26 protons to the apomyoglobin molecule. n+ designates the $(M+nH)^{n+}$ ion, where M is the protein molecule. The flow rate used in this experiment was 0.5 μl/min. Therefore, the sample quantity consumed during the acquisition period was 500 fmol. It is interesting to note that 500 fmol corresponds to 100 nl of protein solution. It is impractical with the present system to inject and electrospray such a small volume of liquid. Therefore, the total sample quantity that can be handled with ease and confidence for a complete analysis including stabilization of spray and signal is generally higher (~10 pmol). Use of small volume injection loops for sample introduction to the spray ionization source together with the use of lower protein concentrations may further reduce the sample quantity required for their analysis.

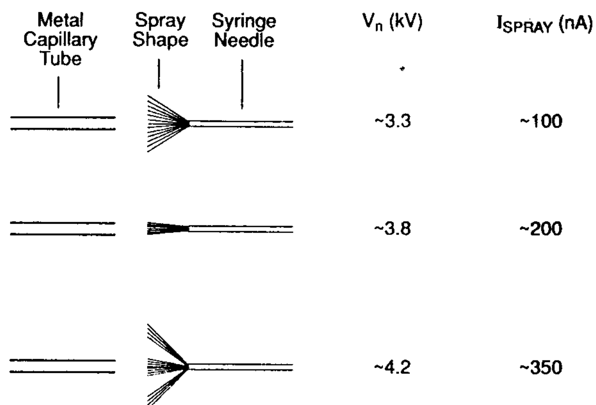

Figure 3 Demonstration of the electrospray shape obtained from protein solutions (1-10 μM) comprised of water: methanol: acetic acid = 45 : 52 : 3 at different voltages, Vn, applied to the syringe needle. I_{spray} denotes the spray current. The distance between the metal capillary tube and the needle was 0.55 cm. Flow rate = 0.5 μl/min. The spray current was measured as shown in Fig. 2.

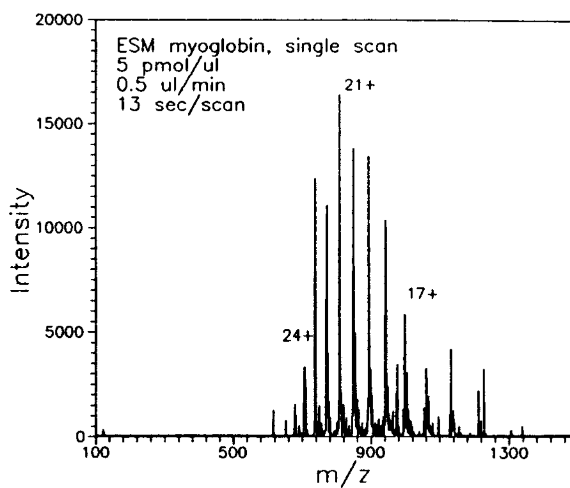

Figure 4 Electrospray ionization mass spectrum of equine skeletal muscle myoglobin. The spectrum is a single scan acquired in 13 s from a 5 μM myoglobin solution (water : methanol : acetic acid = 45 : 52 : 3% v/v). n+ indicates the number of protons attached to the molecule. The flow rate was 0.5 μl/min.

We have investigated ~50 protein samples by ESI mass spectrometry. The proteins are observed to provide a widely different mass spectrometric responses. The spectrum of equine myoglobin shown in Figure 4 represents an example of a protein that produces a strong response. However, many glycoproteins, membrane proteins, bacterial enzymes, and proteins containing multiple intact disulfide bonds were observed to yield lower responses than those obtained for proteins like myoglobin, cytochrome C, and carbonic anhydrase II. Jardine [8] also reports that attempts of analyses of certain proteins by ESI mass spectrometry have not been successful.

Because different proteins provide widely different mass spectrometric responses, the analysis of protein mixtures by ESI mass spectrometry is particularly difficult. Attempts to quantitate protein components of a mixture whose ionization efficiencies are not known can be highly misleading. For example, an human hemoglobin molecule consists of two α-chain and two β-chain proteins non covalently attached to each other [9]. The two chains contain similar number of amino acids (α-chain 141; β-chain 146) and are highly homologous. Because noncovalently attached protein molecules do not remain together after electrospray ionization, it is therefore expected that equal intensity ions from both α- and β- chains will be observed in the elctrospray ionization mass spectrum of hemoglobin. However, the electrospray ionization mass spectrum of normal human hemoglobin (Sigma cat. # H7379), shown in Fig. 5, is dominated by multiply charged ions originating from the α-chain and the ions from the β-chain constitute only 20.8% of the total ion intensities arising from both chains.

The reasons for wide range of ESI mass spectrometric responses obtained with several proteins are not clear. One important factor that may contribute to inefficient ionization is the higher order protein structure [10]. It is possible that under the ionization conditions normally used, the majority of the protein molecules are in a conformation that does not present a large number of basic sites for ready protonation. The resulting lowly charged ions with high m/z values may remain beyond the m/z range of the quadrupole analyzer. Only a small fraction of the protein molecules may be in a conformational state that allows a large number of charges to be added to the molecule. Smith and co-workers [2,11] have shown that reduction of disulfide bonds of proteins result in the formation of unfolded molecules that produce ions with much higher number of charges than that obtained from an unreduced protein. We have noticed that proteins, like, metallothionein containing no disulfide bond also produce very weak responses.

MOLECULAR MASSES

Molecular masses can be readily determined from the observed m/z values of peptide and protein ions in the ESI mass spectra provided that the charge, z, of at least one ion, arising from the analyte of interest, is known. The procedure adopted for the determination of the charge of an ion, when a distribution of ions is observed, has been described in the literature [1,2]. The accuracy of the observed molecular masses depends on the accuracy to which the m/z values are measured. Table 1 lists the experimentally determined and calculated molecular masses of proteins obtained in the authors' laboratory together with the difference between the observed and calculated masses. The molecular masses of proteins were calculated using the data given in the Dayhoff Protein Sequence Database [12]. In a previous communication [3] we reported the results of an analysis of some 13 proteins by the ESI method. Several of these proteins have been reinvestigated. Table 1 contains twenty three new determinations in addition to those reported earlier [3].

In general, for proteins that exhibit strong mass spectrometric responses, the observed molecular masses agree, to within 200 ppm, with the corresponding calculated values

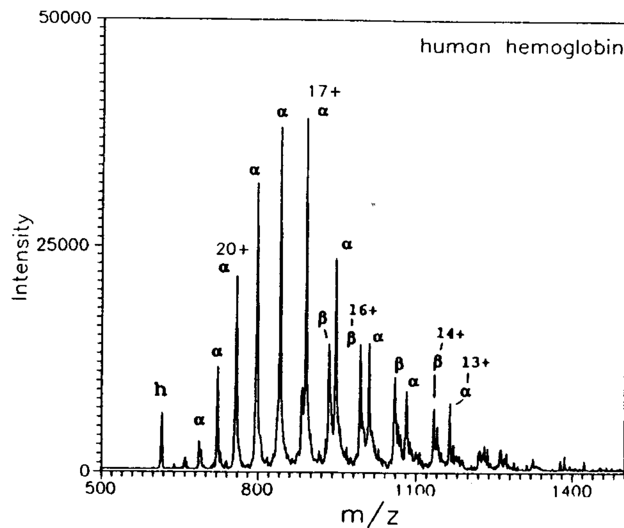

Figure 5 Electrospray ionization mass spectrum of human hemoglobin. Ions arising from the α-chain is designated as α and those from the β-chain as β. h represents an ion from heme. The spectrum is an average of 6 scans (33 s each) obtained from a 20 μM protein solution electrosprayed at a rate of 0.6 μl/min. The solvent composition was the same as in Fig. 4.

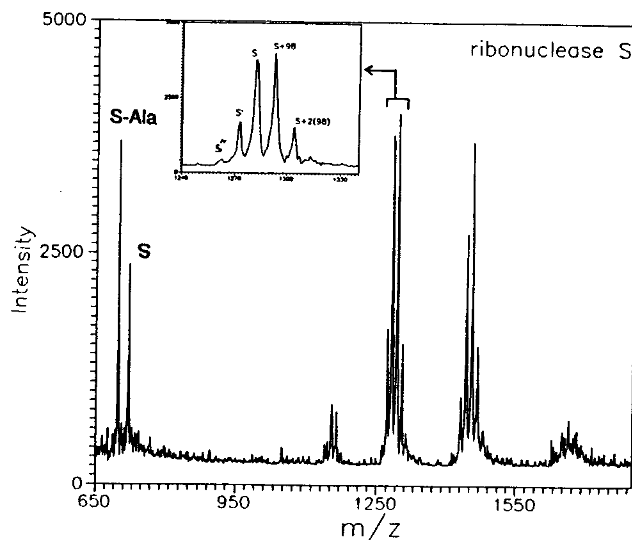

Figure 6 Electrospray ionization mass spectrum of bovine ribonuclease S. The spectrum contains ions of light chain peptides between m/z 650 and 750 and ions from heavy chain proteins between m/z 1000 and 2000. S and S-Ala designate respectively the S-peptide and a peptide without the c-terminal alanine. Inset: detail display between m/z 1240 to 1340. S, S', S'' represent respectively the $(M + 9H)^{9+}$ ions of ribonucease S, ribonuclease S', and ribonuclease S".

Table 1 Observed and calculated molecular masses (MM) of proteins investigated.

Protein	Observed MM(u)	Calculated[a] MM(u)	Δ^b (ppm)
glucagon	3481.3 ± 1.4	3482.4	-316
human β-endorphin	3464.4 ± 0.2	3465.0	-173
bovine insulin B-chain (oxidised)	3495.1 ± 0.7	3496.2	-314
HIV(1-46)Ser[46]	5157.3 ± 1.0	5158.0	-136
bovine insulin	5734.2 ± 0.9	5733.6	+105
bovine ubiquitin (1-24)	8450.0 ± 0.3	8450.9	-107
bovine ubiquitin	8565.0 ± 1.1	8565.0	0
yeast ubiquitin	8557.4 ± 1.4	8556.8	+70
human apolipoprotein C_{III0}	8766.1 ± 0.6	8764.2	+216
bovine cytochrome C	12229.3 ± 1.9	12230.8	-122
equine cytochrome C	12359.1 ± 1.7	12360.9	-145
bovine ribonuclease S	11446.2 ± 1.3	11446.9(S')	-61
	11534.7 ± 1.5	11534.0(S)	+60
	11357.1	11359.8(S")	-239
bovine ribonuclease A	13681.1 ± 1.6	13682.3	-88
chicken egg lysozyme	14308.2 ± 4.2	14305.2	+210
superoxide dismutase	15588.8 ± 1.9	15592.4	-231
bovine nuclease S-7	16805.6 ± 3.2	16807	-83
human hemoglobin			
α-chain	15128.5 ± 1.9	15126.4	+139
β-chain	15869.0 ± 1.6	15868.2	+50
turkey hemoglobin[c]	15338.1 ± 2.3	–	
	16315.4 ± 1.0	–	
	15622.0 ± 1.5	–	
	15773.9 ± 1.0	–	
difolate reductase	17993.3 ± 0.5	17998.4	-284
(wild type)	18159.1 ± 0.8	d	
b-lactoglobulin A	18364.7 ± 1.4	18363.1	+87
trypsin inhibitor	19978.6 ± 0.5	d	
	20091.2 ± 7.1	20091	+10
bovine trypsinogen	23981.6 ± 2.0	23981.1	+20
a-chymotrypsin (bovine)	25235.6 ± 1.6	25234	+64
elastase (porcine)[c]	25966.0 ± 11	–	
concanavalin – A[c]	25608.9 ± 6.4	–	
(jack bean)	12934.5	d	
human apolipoprotein AI	28078.1 ± 0.8	28078.6	-18
proteinase K	28936.3 ± 2.6	28904.9	+1085
carbonic anhydrase II (bovine)	29021.8 ± 1.3	29021.3	+17
subtilisin Carlsburg	27304.0 ± 19	27288.4	+571
	28554.5 ± 1.6	d	
aldolase (recombinant)[c]	32923.0 ± 2.7	–	
bovine serum albumin	66509.0 ± 23	66267	+3650
Glycoproteins			
apolipoprotein C_{III1}	9420.6 ± 1.2	9420.8	-21
apolipoprotein C_{III2}	9712.1 ± 0.8	9712.1	0
turkey egg conalbumin[c]	77563.0 ± 23	–	
bovine transferrin	78275.0 ± 73	–	

a: Molecular masses are calculated using the sequences compiled in the Dayhoff Protein Sequence Database[12].

b: Difference between the observed (column 2) and the calculated molecular mass (column 3).

c: Protein sequence unavailable.

d: A protein species whose molecular mass could not be correlated with the sequence of a known protein variant.

(Table 1). Similar accuracies in MM measurements are also observed by other investigators [1,2]. Three notable exceptions to the above agreements are the observed molecular masses of subtilisin Carlsburg, proteinase K, and bovine serum albumin. The reason(s) for the observed discrepancies in the measured and calculated molecular masses of these proteins are being investigated. These three proteins also exhibited weak mass spectrometric responses and relatively wide peaks. The observation of wide peaks together with higher molecular masses may be a consequence of sample heterogeneity and/or incomplete removal of adduct species, such as, H_2SO_4, H_3PO, etc. [13].

The ESI mass spectrum of turkey hemoglobin (Sigma cat. # H0142) exhibited the presence of 4 protein components (see Table 1). Because the sequences of turkey hemoglobin chains are not available, assignment of the observed molecular masses to appropriate hemoglobin chains is not possible at this time.

Several other protein samples also produced ions corresponding to multiple protein components. Thus, a sample of ribonuclease S (Sigma cat. # R6000) produced ions, in the m/z range 1000-2000 (Fig. 6), from two proteins: ribonuclease S (obs. MM = 11534.7 u, cal. MM = 11534.0 u) and a protein with a MM 87 u lower (Table 1). The latter corresponds to a known S-protein variant, the S'-protein, that lacks the NH_2-terminal serine of ribonuclease S [14]. Weak ions corresponding to a third protein with a MM 87+87 u lower than ribonuclease S is also observed. We designate this previously unobserved varient S", which appears to lack a second serine residue. In the lower m/z region (650-750) two triply protonated ions from two peptides are also observed. The one with m/z 722.9 (designated S) corresponds to the S-peptide (obs. MM = 2165.7 u, cal. MM = 2166.3 u) and the other (designated S-Ala, obs. m/z 699.3) has a MM that is 70.9 u lower than the S-peptide. Ribonuclease S is produced by proteolytic cleavage of ribonuclease A at Ala(20)-Ser(21) resulting in the formation of the S-peptide (1-20) and the S-protein (21-124). The S'-protein is obtained when the hydrolysis occurs at Ser(21)-Ser(22). In the present investigation no corresponding S'-peptide (1-21) was observed. The second peptide observed in Fig. 6 corresponds to residues (1-19). Additional ion peaks, such as S+98, S+2(98), shown in the inset of Figure 6 arise from the adduction of sulfuric acid and/or phosphoric acid to the protonated protein ions [13]. Loo et al. [15] have also investigated ribonuclease S by the ESI technique. However, these authors did not report the observation of ions from the S' or S"-protein.

Multiple components observed with several other proteins, given in Table 1, could not be identified. A significant number of the commercial protein samples used in the present investigation are observed to contain more than one protein. Because proteins yield different electrospray ionization mass spectrometric responses, the impuritiy ions are sometimes observed to dominate the mass spectra and great care is needed in the interpretation of the mass spectra. Despite these difficulties, we have found ESI to be an enormously powerful technique for the analysis of biopolymers.

ACKNOWLEDGEMENT

This work was supported in part by grants RR00862 and RR07065, Division of Research Resources, NIH, and GM38274, National Institute of General Medical Sciences, NIH. We thank Mrs. Gladys McMilleon for her assistance in preparing the manuscript.

REFERENCES

1. J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong, and C.M. Whitehouse, *Science*, **246** (1989) 64.
2. R.D. Smith, J.A. Loo, C.G. Edmonds, C.J. Barinaga, and H.R. Udseth, *Anal. Chem.* **62** (1990) 882 and references therein.
3. S.K. Chowdhury, V. Katta, and B.T. Chait, *Rapid Commun. Mass Spectrom.* **4** (1990) 81.
4. V. Katta, S.K. Chowdhury, and B.T. Chait, *Anal. Chem.* (submitted).
5. S.K. Chowdhury, V. Katta, and B.T. Chait, *Biochem. Biophys. Res. Commun.* **167** (1990) 686.
6. J.B. Anderson, R.P. Andres, and J.B. Fenn., *Adv. Chem. Phys.* **10** (1966) 275.
7. D.M. Moltz, R.A. Gough, M.S. Zisman, D.J. Vieira, H.C. Evans, and J. Cerny, *Nucl. Instr. Methods* **172** (1980) 507.
8. I. Jardine, *Nature* **345** (1990) 747.
9. J. Darnell, H. Lodish, and D. Baltimore, *Molecular Cell Biology*, W.H. Freeman, NY, 1986, pp. 57-9.
10. S.K. Chowdhury, V. Katta, and B.T. Chait, *J. Am. Chem. Soc.* (submitted).
11. J.A. Loo, C.G. Edmonds, H.R. Udseth, and R.D. Smith, *Anal. Chem.* **62** (1990) 693.
12. Protein Identification Resource (PIR), National Biomedical Research Foundation (NBRF) Release 19.0 (3187 December 1988).
13. S.K. Chowdhury, V. Katta, R.C. Beavis, and B.T. Chait, *J. Am. Soc. Mass Spectrom.* **1** (1990) 382.
14. M.S. Doscher and C.H.W. Hirs, *Biochem.* **6** (1967) 304; E. Gross and B. Witkop, *Biochem.* **6** (1967) 745.
15. J.A. Loo, C.G. Edmonds, R.D. Smith, M.P. Lacey, and T. Keough, *Biomed. Environ. Mass Spectrom.* **19** (1990) 286.