ELSEVIER

International Journal of Mass Spectrometry 219 (2002) 39–62

International Journal of Mass Spectrometry

www.elsevier.com/locate/ijms

Origin and number of charges observed on multiply-protonated native proteins produced by ESI

Natalia Felitsyn, Michael Peschke, Paul Kebarle*

Department of Chemistry, University of Alberta, Edmonton, Alta., Canada T6G 2G2

Received 19 July 2001; accepted 30 January 2002

Abstract

Native proteins and particularly native non-covalently bonded protein–protein and protein–substrate complexes are of great interest and are intensely studied by ESI–MS methods. The multiple charges on these ions are not only useful in lowering the *m*/*z* values but play also an important role in the chemical behavior of these complexes.

Evidence from the literature and the present work is presented which supports the charge residue model (CRM) as the mode of formation of the charged globular proteins in the gas phase. Very small water droplets which contain only one protein molecule are ultimately formed in the ESI process. The surface of these droplets is charged by an excess of small ions due to a salt which is also present in the solution. Thus, in the positive ion mode, and when the buffer (ammonium acetate) is the main electrolyte used, the excess small positive ions are NH_4^+ ions. Evaporation of the water in the droplet leads to a residue which is the globular protein. The protein is charged by the excess positive ions such as NH_4^+. The number of NH_4^+ ions available, Z_{CRM}, can be predicted on the basis of CRM. The proteins in order to be able to hold all of the protons provided must have a sufficient number of basic side chains located at the surface of the protein. It is found that most proteins have more than enough basic sites to hold the charge, Z_{CRM}. Examples for these are carbonic anhydrase and cytochrome *c*. For these proteins the charge observed with ESI–MS is found to be close to equal to the charge, Z_{CRM}, supplied.

Some unusual proteins such as pepsin, have too few basic side chains, much less than the number of charges, Z_{CRM}, provided. For these proteins the number of basic sites available on the protein determine how many of the charges provided by CRM will be retained. The number of basic sites can be evaluated and is found in agreement with the observed charges in the mass spectrum.

Other predictions can also be made on the basis of the CRM. Thus, evaporation of the water droplet will lead to formation of neutral (uncharged) adducts on the protein, which are due to neutral components of the buffer. The approximate number of adducts can be predicted. Predictions can also be made which buffers will lead to adducts difficult to get rid off, in the desolvation stage of the mass spectrometer. (Int J Mass Spectrom 219 (2002) 39–62)

Keywords: Native proteins; ESI; Desolvation

1. Introduction

The area involving mass spectrometric studies of non-denatured proteins and protein–protein, protein–substrate, non-covalently bonded complexes, by ESI–MS, has developed very rapidly and is providing unique information of great significance [1]. Non-denatured proteins and complexes are required in all studies which attempt to provide correlations between binding energies of the complexes observed in the gas phase and the binding energies in the biological environment such as the aqueous solution of the

* Corresponding author. E-mail: paul.kebarle@ualberta.ca

1387-3806/02/$ – see front matter
PII S1387-3806(02)00588-2

cytoplasm. The experimental conditions required for obtaining good, clean mass spectra at neutral pH and in aqueous solution, were developed essentially by trial and error. Thus, it was found in early work that the traditionally used buffers in protein biochemistry were generally unsuitable for ESI–MS, because they led to formation of multiple adducts to the protein, such as multiple catonization by Na^+, K^+ ions when present in the buffer or neutral adducts due to neutralized components of the buffer. These obscured the mass spectrum, which desirably should contain only the polyprotonated protein, see Smith and coworkers [2] and references therein.

In present practice [2], the most often used buffer is ammonium acetate (NH_4Ac), while ammonium bicarbonate is also used but to a lesser extent. These buffers generally lead to relatively clean multiply protonated proteins. It is worthwhile to recall that NH_4Ac is not a buffer. The dissolved salt in water does lead to a pH = 6.8 which is very close to the required physiological pH, but NH_4Ac alone in solution has limited ability to buffer, i.e., to resist changes of pH. One major theme of the present work will be to explore the reasons for the success of NH_4Ac and ammonium bicarbonate. Evidence will be given that the major role of these buffers is to provide NH_4^+ ions which are the protonating reagents that lead to the observed multiple protonation of the globular proteins. The recognition of this fact is of considerable practical utility.

To obtain a good understanding of the conditions that lead to the multiple protonation and multiple charging in general, one needs to consider the mechanism which leads to formation of multiply protonated globular proteins in the gas phase.

The mechanism of ESI [3–18] involves three stages:

(a) Formation of small charged droplets at the tip of the electrospray capillary. The droplets are charged because, due to the action of the applied electric field on the solution at the ES capillary tip, the meniscus of the liquid at the capillary tip becomes enriched on positive ions. Therefore, the droplets formed from the elongated meniscus (Taylor cone) contain an excess of positive electrolyte ions, when the capillary tip is positive (positive ion mode).

(b) Evaporation of solvent from the droplet decreases the radius of the droplets and since the charge is conserved, at some critical radius given by the Rayleigh equation [3], Coulombic forces overcome the surface tension of the liquid and lead to fission of the droplets. Repeated evaporation and fission lead to very small charged droplets which are the precursors of the gas phase ions.

(c) The actual mechanism by which the gaseous ions ultimately detected with the mass spectrometer, are formed from the very small and highly charged droplets was until recently under dispute. Two models were considered, the charged residue mechanism (CRM) and the ion evaporation mechanism (IEM). Fortunately recently strong evidence has accumulated that indicates that both mechanisms apply. The IEM is involved in the production of small ions such as the conventional inorganic and organic ions [8,10,14], while CRM is involved in the production of macroions such as the globular proteins [2d,11,18].

1.1. CRM

The charged residue model is due to Dole and coworkers [16]. It focuses on the very last stages of solvent evaporation from the smallest droplets produced by solvent evaporation and successive Coulomb fissions of parent droplets. Some of these very small droplets may contain only one or a few charges, i.e., a few unpaired electrolyte ions. Continued solvent evaporation of such a droplet to dryness will lead to gas phase ions. If the droplet happens to contain also one neutral solute molecule, adduct formation between the ion(s) and the molecule will lead to a charged ion–molecule complex. A cartoon illustrating the formation of a droplet which contains one globular protein as a solute is shown in Fig. 1.

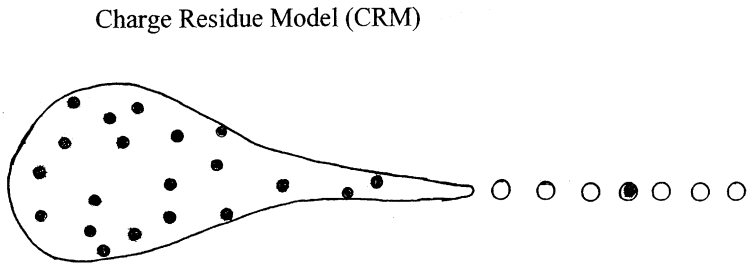


Fig. 1. A cartoon illustrating the production of globular proteins in the gas phase by CRM. A parent droplet with many charges on its surface (charges not shown) undergoes a Coulomb fission. Some of the off-spring droplets contain one globular protein (black disk). Evaporation of the water from such an off-spring droplet leads to a protein in the gas phase. Charges on the surface of the off-spring droplet are transferred to the protein. (Based on Fig. 13 in Kebarle and Ho [18].)

1.2. IEM

The formation of gas phase ions by IEM was first proposed by Iribarne and Thomson [17]. These authors, on the basis of a derived equation, showed that when the charged droplets become very small (radius $R < 10$ nm), the electric field due to the charges at the surface of the droplet is strong enough to cause emission (evaporation) of gas phase ions from the droplet. For example, assuming that the major electrolyte in the solution was NaCl, in the positive ion mode of ESI, the charges at the surface on the droplets will be due mostly to Na^+ ions, while the interior of the droplets will be populated by paired Na^+ and Cl^- ions. The emitted ions are generally partially solvated [17]. Thus, $Na^+(H_2O)_m$, where $m = 6$ or 7, may be expected to evaporate, because the activation energy required for such an evaporation is lower than that for the naked ion [8,17].

1.3. Charged globular protein ions are produced by CRM

The derivation of the IEM equation [17] was based on small ions and it is most doubtful that it can be extended to very big ions like the globular proteins. Furthermore, globular proteins, such as enzymes, are very often designed to function within the cytoplasm and therefore must be soluble in water. This is achieved by the presence of polar (hydrophilic) side chains and ionizable acidic and basic side chains, which are located at the surface of the proteins. Due to the presence of these groups the protein is not likely to seek the droplet surface but will in general tend to stay inside the droplet and be available to become a gas phase ion via the charge residue model.

Another reason to expect CRM for globular proteins, is the fact that the observed proteins are not only charged but also, in general heavily covered with neutral adducts due to the presence of other solutes, such as buffers. These must be removed by heating in a “desolvation” process. The presence of such multiple adduction would be expected with CRM but much less so with IEM.

Globular proteins in the mega-Dalton mass range are routinely produced by ESI. The production by a mechanism where these giant species “evaporate” from the droplet appears very unlikely.

Strong support for the CRM being the mechanism was provided recently by Fernandez de la Mora [11]. He used data compiled by Smith and coworkers [19a], which are shown in Fig. 2. The experimental points give the range of the observed numbers of elementary charges, Z_{obs}, on globular proteins, electrosprayed under conditions which did not lead to denaturing, i.e., pH = 7, and water as the only major component of the solvent. The authors [19a] found that the average Z_{obs} plotted vs. the molecular mass M of the protein (Da) was well fitted by an equation of the form:

$$Z_{obs} = aM^b \quad (1)$$

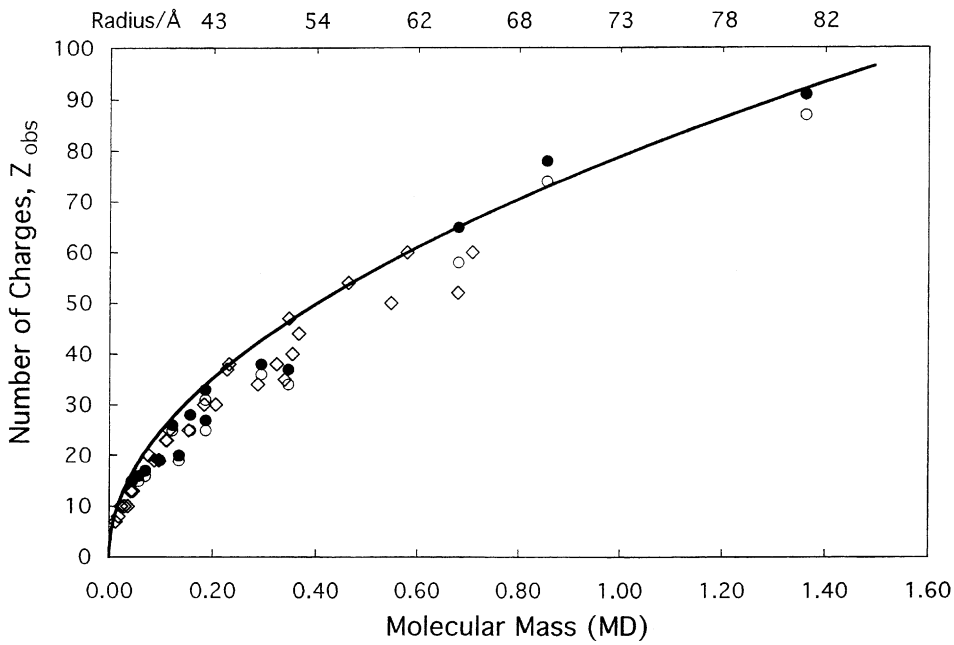


Fig. 2. Reproduction of plot used by Fernandez de la Mora [11]. Number of charges observed mass spectrometrically, Z_{obs}, (●) highest; (○) lowest charge sprayed under non-denaturing conditions, from data in the literature [19a] vs. the mass of the protein (in MDa). Fernandez de la Mora [11] showed, see solid curve, that Z_{obs} is very close to charge Z_R expected on water droplets of the same size as the protein. Also shown, see upper horizontal axis, is the radius for the protein assumed by Fernandez de la Mora. Also included in the present plot are the data obtained by Standing and coworkers (◇) [19b].

where both a and b are constants. The value $b = 0.53$ led to the best fit [19a]. Standing and coworkers [19b] have also provided a plot which involves only non-covalently bonded protein complexes. This plot also provides a value of b which is between 0.52 and 0.55. The data in both plots, extend from $M = 10^4$ to $M = 10^7$ Da, an impressively wide mass range! The data of Standing were included in the plot shown in Fig. 2.

Fernandez de la Mora [11] found that the Z values for the globular proteins could be also well fitted by the charge Z_R predicted by the Rayleigh stability limit for Coulomb fissions [3], of *water* droplets of the same radius R as the globular proteins.

The Rayleigh equation [3] is

$$Z_R = \frac{8\pi}{e}(\gamma \varepsilon_0 R^3)^{1/2} \tag{2}$$

where γ is the surface tension (0.072 N/m for water), ε_0 the electrical permittivity of vacuum (8.8×10^{-12} C^2/J m), e the elementary charge (1.6×10^{-19} C).

Fernandez de la Mora [11] evaluated the radius of the proteins from the molecular mass M of the proteins in Daltons, by assuming that the density of the globular proteins is the same as that of water $\rho = 1$ g/cm^3. (This assumption, which is approximately true, will be examined later.) With this assumption, the radius of the proteins was evaluated with

$$\left(\frac{4}{3}\pi R^3 \rho\right) N_A = M \tag{3}$$

where N_A is the Avogadro number (6×10^{23}).

Combining Eqs. (2) and (3), one obtains

$$Z_R = \frac{\gamma\pi}{e}\left(\frac{\rho\gamma\varepsilon_0}{4\pi N_A}\right)^{1/2} M^{1/2} = 0.078 M^{1/2} \tag{4}$$

(numerical factor 0.078 is for M (Da)).

Notably, the exponent for M equal to 0.5 is very close to the exponent 0.53 of the empirical Eq. (1) observed by Smith and coworkers, and Standing and coworkers [19]. The curve corresponding to Eq. (4) is shown in Fig. 2. Also shown on the horizontal axis are the corresponding values of the radius R, see Eq. (2), corresponding to the given M of the protein, Eq. (3).

Fernandez de la Mora [11] proposed that the observed agreement between the charges on the proteins and the charges on water droplets of the same size, at the Rayleigh limit, is not a coincidence but a consequence of the multiply charged proteins being formed by the CRM. Small droplets just a little bit larger than the protein will have a (excess) charge at the surface, close to that at the Rayleigh limit [11]. In some of these droplets one protein molecule could be present. As the solvent evaporates, and the charged water surface collapses over the protein, the (excess) charges, due to the small ions at the surface, will be transferred to the protein, and for this reason one can expect the charge on the gaseous protein Z_P will be close to the Rayleigh charge on the water droplet, Z_R.

Fernandez de la Mora's work [11] was preceded by a very similar proposal due to Smith et al. [2d] which did not deal with proteins but with starburst dendrimers. However, Smith's work provided much less details and documentation.

1.4. The chemistry of charging of globular proteins

Fernandez de la Mora [11] considered in some detail various aspects of the production of the charged gas phase proteins. However, the question as to the chemical nature of the charges on the droplet surface and its relationship to the chemical nature of the charges on the proteins and specifically the formation of polyprotonated proteins (in the positive ion mode) was not examined.

The experience, in modern work on protonated globular proteins at neutral pH, that ammonium acetate is the best buffer, derives in our view, not so much from the quality of ammonium acetate as a buffer, but from the ability of NH_4^+ to protonate the basic side chains at the surface of the protein. The ability of NH_4^+ to protonate basic side chains has been proposed earlier by Siu and coworkers [20a] in a very interesting study of the protonation of a small protein, gramicidin S. NH_4^+ as protonating agent was considered also in more recent work [18,20b].

The concentrations typically used in modern work on non-denatured proteins are found to be $\sim 10^{-5}$ M in the globular proteins and 10^{-3} M or higher in ammonium acetate. Thus, the concentration of NH_4^+ is in general at least 100 times higher than that of the protein. This fact shows that NH_4^+ is expected to be the major charged constituent at the surface of the droplets.

A schematic representation of the charging of the globular protein by NH_4^+ ions at the surface of the droplets is shown in Fig. 3. As the last water evaporates and assuming a near neutral pH, the basic and acidic side chains on the surface of the proteins, which were ionized in solution, will become neutralized by reacting predominantly with the major counter ions in the vanishing aqueous solution:

$$\text{protein}\cdots NH_3^+ + Ac^-(aq) = \text{protein}\cdots NH_2HAc\ (\text{neutralization}) \quad (5)$$

$$\text{protein}\cdots CO_2^- + NH_4^+(aq) = \text{protein}\cdots CO_2HNH_3\ (\text{neutralization}) \quad (6)$$

The basic side chain in Eqs. (5),(7),(8) is assumed to have an amino group. When the solvent is reduced by evaporation to a (few) monolayers over the protein, near gas phase condition prevail and charging of the basic groups at the surface of the protein, by the NH_4^+ charges at the surface of the droplet can occur. The most stable products will be the proton bridged adducts:

$$\text{protein}\cdots NH_2 + NH_4^+ = \text{protein}\cdots NH_2\text{–}H\text{–}NH_3^+ \quad (7)$$

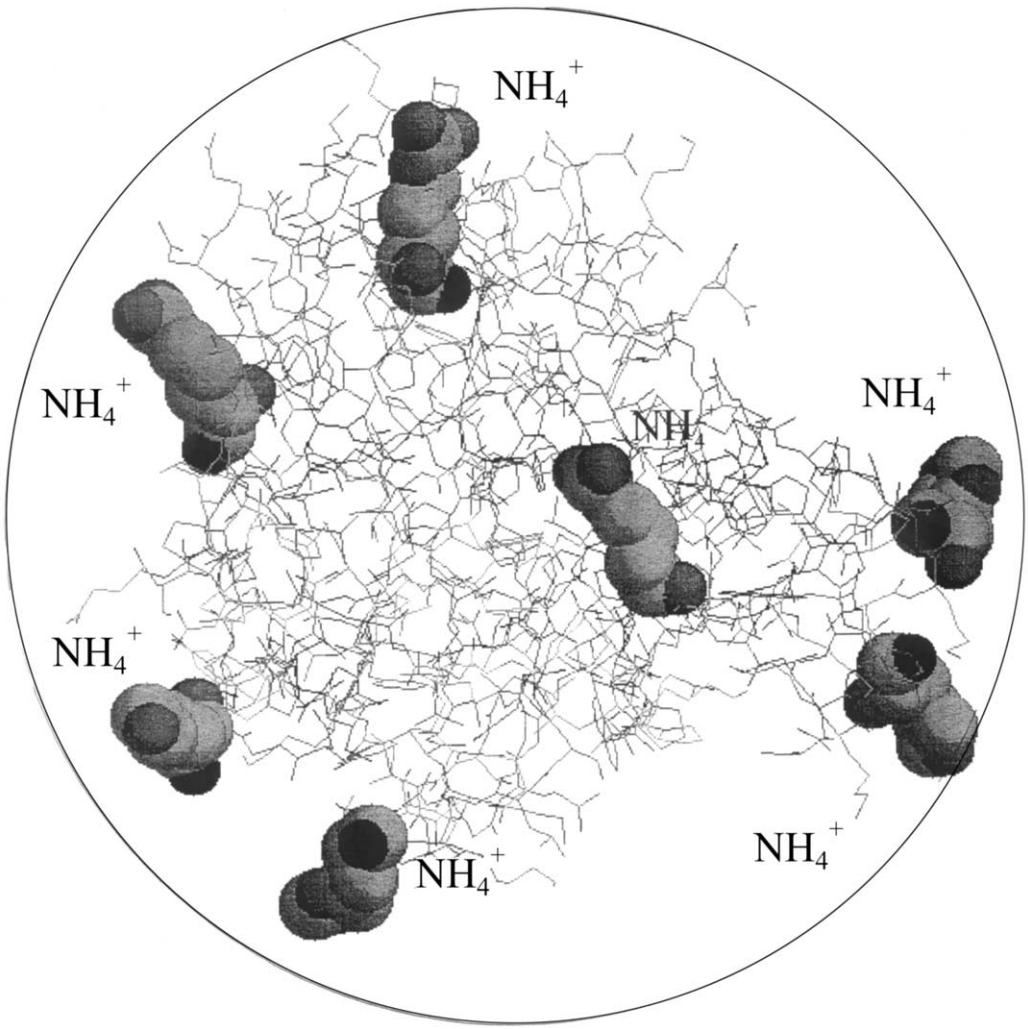


Fig. 3. Schematic representation of carbonic anhydrase, CAII (based on crystallographic data) inside a water droplet. The basic side chains Lys, Arg, His, at the frontal face of CAII are highlighted. NH_4^+ ions present at the surface of the shrinking water droplet, which are part of the excess charge on the droplet, are expected to attach themselves to the basic sites and ultimately lead to their protonation.

A complete proton transfer, leading to protonation of the side chain:

$$\text{protein} \cdots \text{NH}_2\text{–H–NH}_3^+ \rightarrow \text{protein} \cdots \text{NH}_3^+ + \text{NH}_3 \quad (8a)$$

occurs probably later in the "desolvation" stage in the sampling system, either in the heated sampling capillary leading to the mass spectrometer or in the CAD stage due to ion acceleration by the electric field applied between sampling capillary and skimmer electrodes. Details considering processes (7) and (8), are considered in the Section 3. In particular, it will be shown that the energy supplied to achieve desolvation may also lead to charge loss, i.e., the process (8a) is not the only possible outcome. Charge loss via process (8b), which is driven by the Coulombic repulsion

between the other charges on the protein and the leaving charged group can also occur:

$$\text{protein}\cdots\text{NH}_2\text{–H–NH}_3{}^+ \rightarrow \text{protein}\cdots\text{NH}_2 + \text{NH}_4{}^+ \quad (8b)$$

This means that the observed charge Z_{obs}, in the mass spectrum will not depend only on the charging via the CRM mechanism but also on the ability of the protein to hold the charge when processed through the "desolvation" stage.

Experiments which support the proposed mechanism of protonation of globular proteins by the $NH_4{}^+$ ions are presented in Sections 1–3. Charge loss is considered in Section 2.

Another empirical reason for the choice of NH_4Ac was that this salt does not lead to protein ions which include many neutral adducts (such as NH_4Ac or HAc) [2]. Such adducts would be expected with any buffer because of the large molar ratio of buffer to protein used. As already mentioned this is an expected consequence of CRM. Evidently such adducts should form also with NH_4Ac, but are "volatile" enough to be removed in the desolvation process. The HAc and NH_3 adducts, resulting from the neutralization reactions, Eqs. (5) and (6), are also expected to be easily removed in the desolvation process. A brief examination of these processes is included in Section 3.

2. Experimental

Mass spectra were obtained using an Apex II 47e Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer (Bruker, Billerica) with an external electrospray source (Analytica, Branford). Nanoelectrospray was performed using an aluminosilicate capillary (0.68 mm i.d.), pulled to approximately 20 μm o.d. and 1–5 μm i.d. at one end using a micropipette puller (Sutter Instrument Co.). The nanospray tips were positioned approximately 1 mm from the sample capillary using a micro-electrode holder (Warner Instrument Inc.). The electric field required to spray the solution was established by applying a voltage of 800–1000 V to a platinum wire inserted inside the glass tip. The solution flow rate ranged from 5 to 60 nL/min depending on the diameter of the nanoelectrospray tip, electrospray voltage and composition of the solution. The droplets formed at atmospheric pressure were introduced into the vacuum system of the mass spectrometer through a stainless steel capillary (0.43 mm i.d.) operated at a temperature of 150 °C to assist with solvent evaporation from the droplets and microsolvated complexes. The ion/gas jet escaping from the capillary (52–54 V) was transmitted through a skimmer (2–6 V, 1.0 mm i.d.) and stored, electrostatically, in the hexapole. Ions were accumulated in the hexapole for 5–10 s, depending on the ion intensities, then ejected and accelerated to ~2700 V through the fringing field of the 4.7 T magnet, decelerated and introduced into the ion cell. The trapping plates of the cell were maintained at a constant potential of 1.3 V throughout the experiments. The typical base pressure for the instrument was 3×10^{-10} mbar.

Data acquisition was controlled by an SGI R5000 computer running the Bruker Daltonics XMASS software, version 5.0. The time domain spectra, consisting of the sum of 10–40 transients containing 128 K data points per transient, was subjected to one zero-fill prior to Fourier transformation.

All protein samples were purchased from Sigma, and used without further purification. Stock solutions containing 10–20 mg/mL of protein were dissolved in deionized water and refrigerated between use at −20 °C. On the day of the experiment they were diluted to final concentration and buffer, and the other reagents (NaCl, KCl, RbCl or CsI) were added immediately prior to the experiment. At that time pH was measured with Orion 710 model pH meter and Orion micro-pH electrode. The intensities observed for the protein peaks were sensitive to the source conditions: nanospray tip dimensions, solution flow rate, temperature of the metal capillary, nanospray voltage, source voltages (capillary, skimmer and hexapole) and hexapole accumulation times. For the experiments where comparable conditions were desirable, Figs. 4–7, 9 and 11, care was taken to change conditions as little as possible.

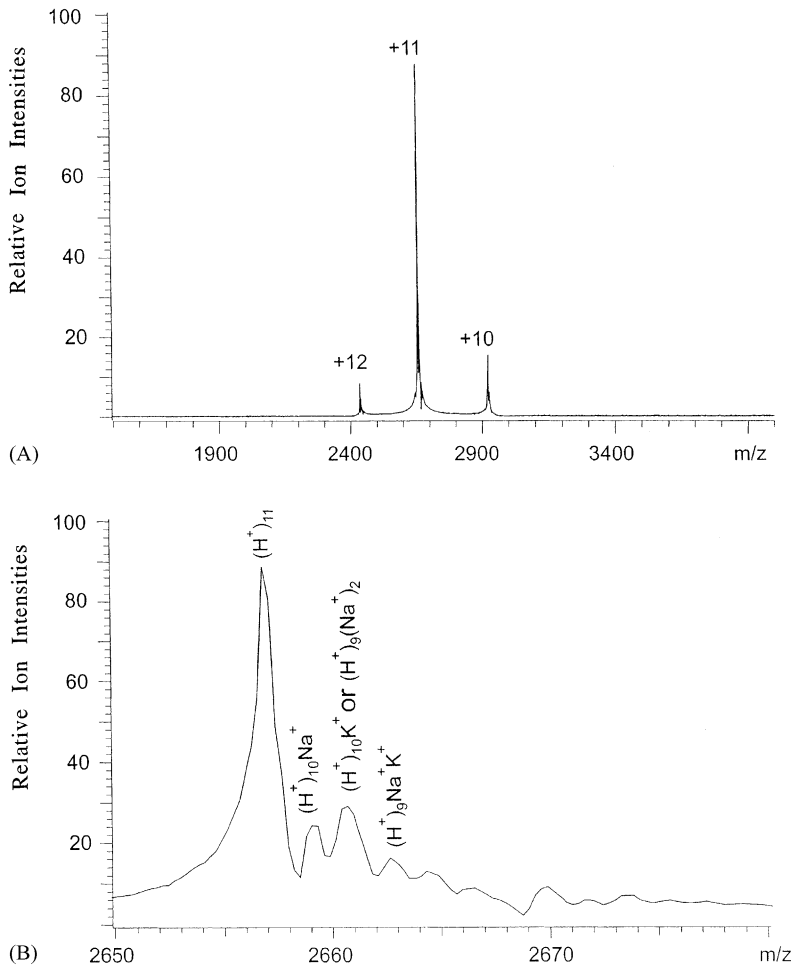


Fig. 4. (A) Mass spectrum observed with 10^{-5} M CAII and 10^{-3} M NH_4Ac, pH = 6.8, in aqueous solution. Major charge state observed $Z_{obs} = 11$. (B) Expanded view of $Z = 11$ peak shows the fully protonated $(H^+)_{11}$ species to be the major ion. However, cationized adducts due to sodium and potassium contamination of the CAII sample are also present.

3. Results and discussion

3.1. Ammonium acetate as protonating agent of proteins at neutral pH

3.1.1. Carbonic anhydrase

To examine the effect of NH_4Ac on the protonation of globular proteins, aqueous solution with a concentration of $\sim 10^{-5}$ M of the protein was electrosprayed with NH_4Ac, at several NH_4Ac concentrations. Human carbonic anhydrase (CAII), (MW 29,180, which includes 64 Da for the Zn cofactor) [21], was sprayed in 2, 1, 0.3 and 0.1 mM aqueous solution of NH_4Ac. CAII was chosen because it is a fair-sized protein [21] with close to spherical shape and a radius of ~23 Å, which combined with the molecular weight leads to a density of 0.96 g/cm^3. The pH of all of the above solutions was 6.8.

Some of the results, obtained with the FT-ICR instrument and nanospray, see Section 2, are given for 1 mM NH_4Ac in Fig. 4, and 0.3 mM NH_4Ac in Fig. 5.

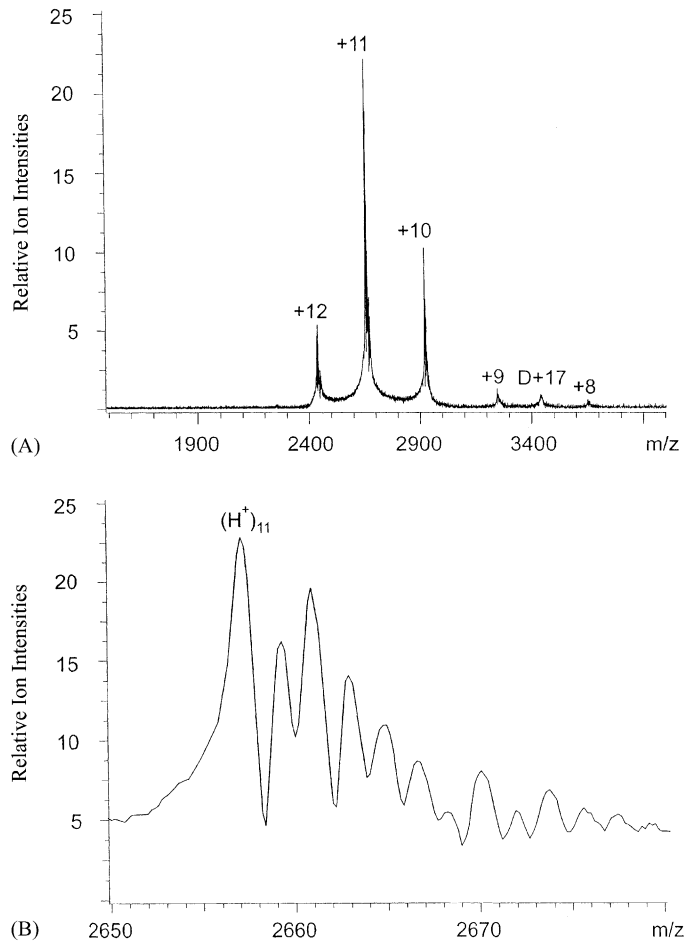


Fig. 5. (A) Mass spectrum of CAII under otherwise the same conditions as in Fig. 4, but NH_4Ac concentration is only 0.3×10^{-3} M. (B) Expanded spectrum of Z = 11 peak shows that lower NH_4^+ to impurities ratio leads to lower degree of protonation relative to cationization due to impurities. The small peak, D + 17, upper spectrum, is due to a small amount of an accidental two CAII molecules complex with 17 charges originating from a larger droplet. The two adjacent peaks indicated as +9 and +8 are also probably due to dimers with +18 and +16 charges.

It should be noted that the relative intensities of the charged states observed under non-denaturing spray conditions, do depend to a certain extent on the specific conditions of the measurements. Thus, Standing and coworkers [22] did observe the $Z = 9$, 10, 11 states for CAII but with $Z = 10$ rather than $Z = 11$, being of maximum intensity, see Fig. 4A. These workers also used ammonium acetate, but at a 6 mM concentration, and electrospray rather than nanospray.

The results for the expanded spectra of $Z = 11$, Figs. 4B and 5B, illustrate clearly the much cleaner protonation for 1 mM NH_4Ac relative to 0.3 mM NH_4Ac. The fully protonated species, $(CAII)H_{11}{}^{+11}$ is the dominant peak accounting for some 50% of the total $Z = 11$ intensity for the1 mM solution. For the 0.3 mM NH_4Ac solution this intensity is down to only 18%, see Fig. 5B. Replacement of H^+ with Na^+ and K^+ are the major source of cationization in the $(CAII)H_{11}{}^{+11}$ spectrum, Fig. 5B. The source of Na^+ and K^+ is likely salts of these ions, which are present in the commercial CAII sample. The intensities in the spectrum, Fig. 4B, indicate that there is more K^+

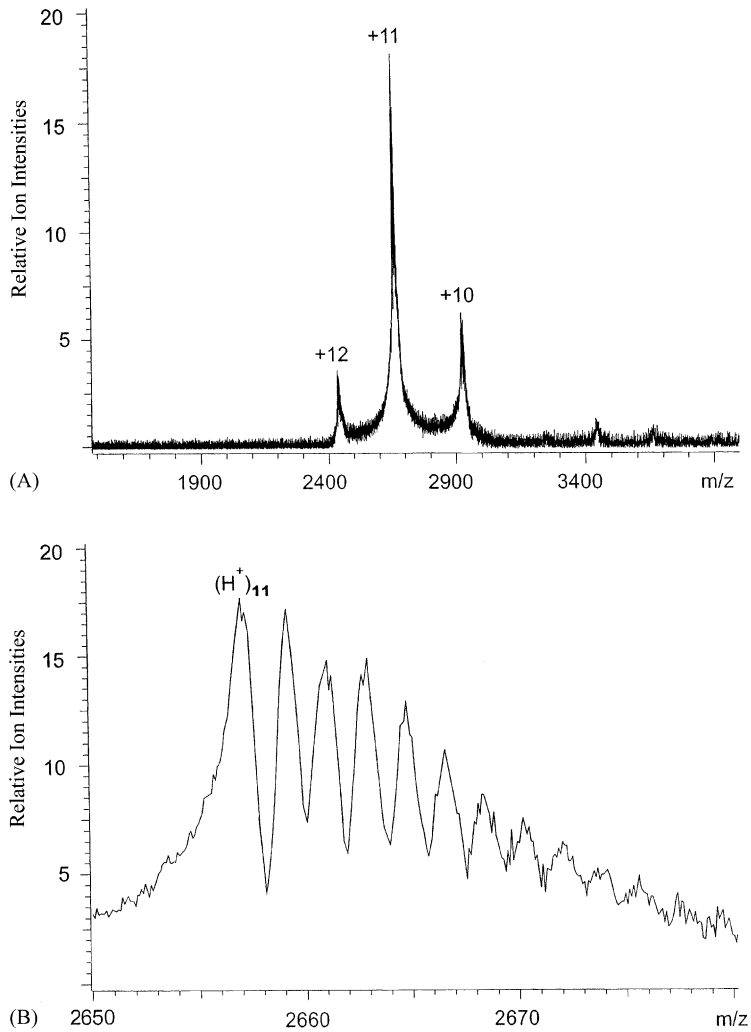


Fig. 6. (A) Mass spectrum from CAII at 10^{-5} M concentration NH_4Ac 10^{-3} M and added NaCl 0.125×10^{-3} M. (B) Expanded spectrum of $Z = 11$ charge state shows extensive sodiation due to added Na^+.

than Na^+ since the $(CAII)H_{10}Na^{+11}$ peak is lower than that of the $(CAII)H_{10}K^{+11}$. However, the observed $(CAII)H_{10}K^{+11}$ has a small contribution from partial overlap with an expected $(CAII)H_9Na_2{}^{+11}$ peak which occurs only ~6 *m/z* units higher and is not resolved from the $(CAII)H_{10}K^{+11}$ peak. The mass spectrum observed with 0.1 mM NH_4Ac (not shown) was much more cationized than the 0.3 mM spectrum, confirming the trend observed in Figs. 4 and 5.

Another small but noticeable effect, is the extension of charged states towards lower Z values, when the concentration ratio: (alkali salt impurity)/(NH_4Ac) increases. Thus, at low cationization by Na^+, K^+ (Fig. 4A), $Z = 10, 9, 8, 7$ charge states are of very low intensity, while at the high cationization, Fig. 5A, these lower charge states have somewhat increased intensities.

The competition between $NH_4{}^+$-induced protonation and cationization by alkali ions was made even

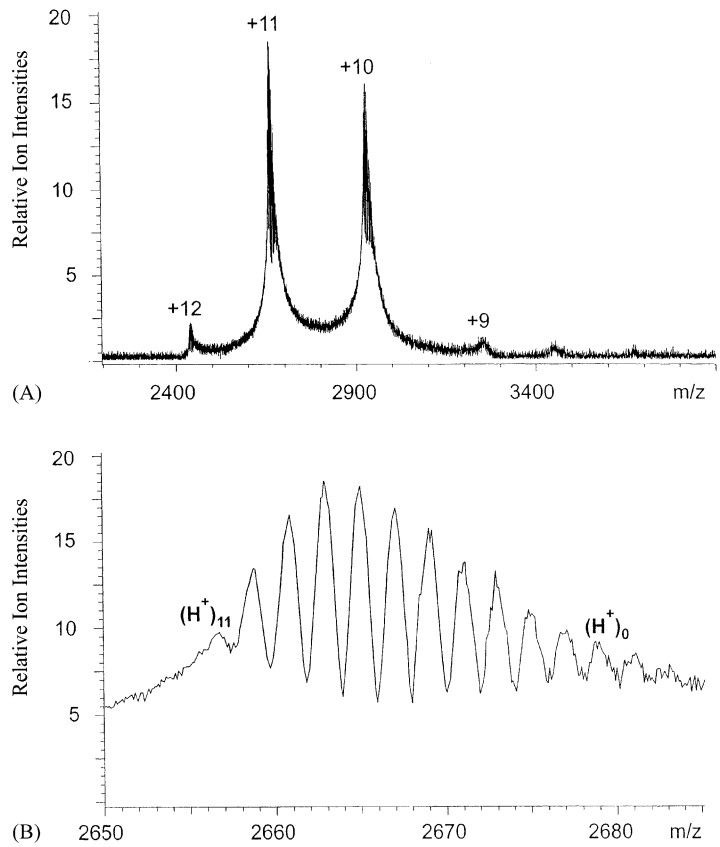


Fig. 7. (A) Mass spectrum of CAII 10^{-5} M, NH_4Ac 0.5×10^{-3} M and NaCl 0.5×10^{-3} M in aqueous solution, pH = 6.8. (B) Extensive sodiation is obtained, see expanded spectrum of $Z = 11$. $(H^+)_0$ corresponds to $(Na^+)_{11}$. A slight shift to lower charge states, compare with Fig. 4, is also observed.

more evident in a series of experiments, where known concentrations of salts like NaCl or other alkali halides MCl, were added to the solution. Mass spectra obtained with 1 mM NH_4Ac, 0.125 mM NaCl are shown in Fig. 6 and 0.5 mM NH_4Ac, 0.5 mM NaCl in Fig. 7. It is evident that an increase of the NaCl/NH_4Ac ratio leads to increased sodiation relative to protonation. This was qualitatively confirmed also by mass spectra (not shown) with other concentration ratios.

The observed intensities of the $(CAII)H_xNa_{11-x}{}^{+11}$ peaks obtained with the 0.5 mM NH_4Ac, 0.5 mM NaCl solutions (Fig. 7B) were used to evaluate the average Na^+ and H^+ content of the $Z = 11$ ions. The result obtained is 46% Na^+ and 54% H^+. This is very close to the 50% NH_4Ac, 50% NaCl present in the solution. However, such close correspondence was observed only in the 1:1 range. Thus, for a Na^+:$NH_4{}^+$ ratio of 1:4 in the solution the ratio in the mass spectrum was 1:2.6, a relative enrichment of Na^+ at low Na^+:$NH_4{}^+$ ratios.

The $(CAII)H_xNa_{11-x}{}^{+11}$ ion intensities (Fig. 7B) are shown in Fig. 8, together with relative intensities from statistical probabilities evaluated with the binomial equation:

$$(p+q)^{11} = p^{11} + 11p^{10}q^1 + 55p^9q^2 + \cdots + q^{11} \quad (9)$$

where $p = 0.54$ and $q = 0.46$. The predicted statistical intensities were evaluated with Eq. (9) by

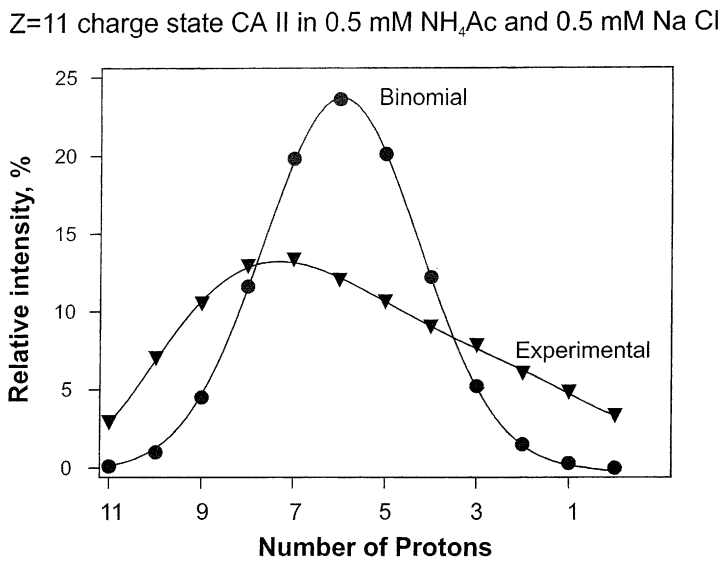


Fig. 8. Plot of intensities of peaks for charge state $Z = 11$ corresponding to $(CAII)H_xNa_{11-x}{}^{+11}$ from an average of three mass spectra including Fig. 7B vs. calculated intensities expected on the basis of a statistical distribution, see Eq. (9).

multiplying the values of the individual terms on the right side of Eq. (9) with the total intensity observed for the $(CAII)H_xNa_{x-1}{}^{+11}$ ions.

The two intensity distributions are quite different, which is not surprising considering that the Na^+ and H^+ ions are expected to be bound to different sites. The H^+ to the basic side chains and the Na^+ to peptide carbonyl oxygens of the back bone as well as to side chains expected to bond strongly to Na^+ in the gas phase. The availability of such different sites and different bonding is expected to skew the random distribution assumed by Eq. (9).

A rough estimate of the concentration of Na, K salts present by weight in the commercial sample of CAII, can be obtained by comparing mass spectra obtained with 1 mM NH_4Ac and 10^{-5} M CAII (see for example Fig. 4B) with the spectra obtained when a known amount of contaminant such as 0.125 mM NaCl was added to the 1 mM NH_4Ac, 10^{-5} M CAII solution (Fig. 6B). Comparison of the observed intensities of cationized peaks indicates that the spectrum Fig. 4B, contains some 0.05 mM Na^+, K^+ solution, which corresponds to a NaCl, KCl content of 1–2% in the commercial CAII sample.

Experiments with MX salts of the alkali ions K^+, Rb^+, Cs^+ were also performed using mixtures of 0.5 mM NH_4Ac and 0.5 mM MX. The mass spectra obtained were, in some cases, more complex than those observed with NaCl. This complexity was due to the presence of neutral adducts. For example, the $Z = 10$, observed with KCl, contained at least one KCl adduct.

An overview, Fig. 9, shows the spectra for 1 mM NH_4Ac, without MX and the spectra of mixtures: 0.5 mM NH_4Ac, 0.5 mM MX (M = Na, K, Rb, Cs). The peak groups for a given charge state show the expected broadening in the series H^+, Na^+, K^+, Rb^+, Cs^+. However, a decrease of the charge Z is also observed in the same order. When essentially only protonation (1 mM NH_4Ac alone) is present, the intensities of $Z = 12$ and 11 (maximum) are the highest, while for Cs^+, the $Z = 12$ and 11, groups have essentially disappeared. Although the shift to lower charge states is evident, it is not dramatic. Larger shifts in the series, Na–Cs, would probably be present in the complete absence of protonation, i.e., in the absence of NH_4Ac. As already mentioned, the positions of attachment of M^+ ions are probably the carbonyl

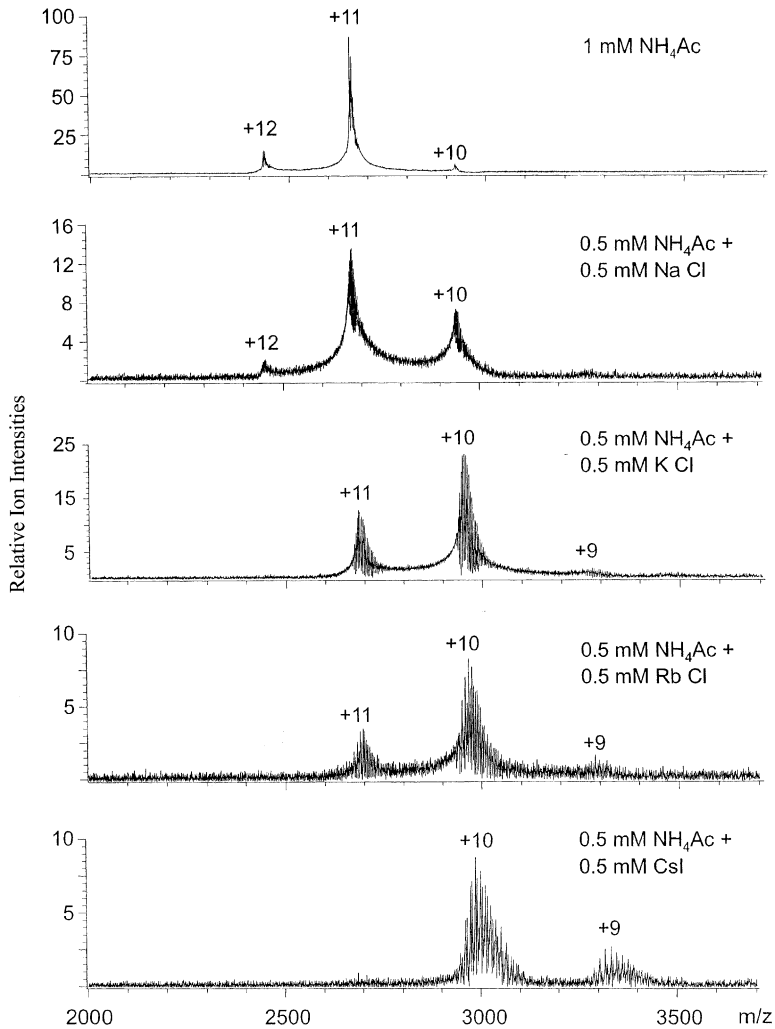


Fig. 9. Mass spectra observed with CAII 10^{-5} M, NH_4Ac 0.5×10^{-3} M and 0.5×10^{-3} M of MX = NaCl, KCl, RbCl and CsI. These are compared with spectrum with no MX addition and 10^{-3} M NH_4Ac shown at the top of the figure. Introduction of cationization by $Na^+ \rightarrow Cs^+$ leads to a small decrease of the charge state from $Z = 11$ for no cationization, to $Z \approx 9.8$ for Cs^+.

oxygens on the peptide back bone or on side chains with carbonyl oxygens. The behavior of such complexes in the cleanup stage, will be quite different, see Section 3.1.1, from that of the ammoniated protons on the basic side chains. The shift to lower charge states may be due to a greater facility for loss of M^+ ions in the desolvation stage. This question deserves a separate study involving addition of MX only, although one would not be certain that the protein does not denature partly, under such conditions.

3.1.2. Cytochrome c

Investigations along the same lines as for CAII were performed also with bovine cytochrome *c* (CYC). This is a considerably smaller protein (MW 12,358, including the heme group). CYC is only roughly spherical, with an approximate radius of 17 Å. Its structure is also unusual in-so-far that there is very little α-helix and β-sheet. In essence, the polypeptide chain of 10^4 residues is wrapped around the heme with residue 1–47 on one side of the heme group and 48–91 on the

other side. As is usual for proteins that must function in aqueous medium, the hydrophobic side chains are on the inside and the hydrophilic side chains are on the outside of the protein. These are clusters of lysine side chains at the crevice along the plane of the heme group [21,23].

The mass spectra obtained with nanospray and the FT-ICR spectrometer (see Section 2) are shown in Fig. 10A and B with solutions of $\sim 10^{-5}$ M CYC and 1 mM NH_4Ac in water. Essentially only two charge states $Z = 8$ and $Z = 7$ are observed. The expanded spectrum of $Z = 8$, Fig. 10B, shows very small intensities of the Na^+ and K^+ adducts. Similar intensity distribution was observed also for the $Z = 7$ state. The low level of Na^+, K^+ adducts may have been assumed to be due to the commercial CYC sample having less Na, K salt contaminants than the CAII sample.

The effect of adding NaCl is shown in Fig. 11 which was obtained with 0.5 mM NH_4Ac and 0.5 mM NaCl. The Na^+ adducts are clearly observed, however the extent of adduct formation is very much lower than was the case for CAII for the same NH_4Ac and NaCl

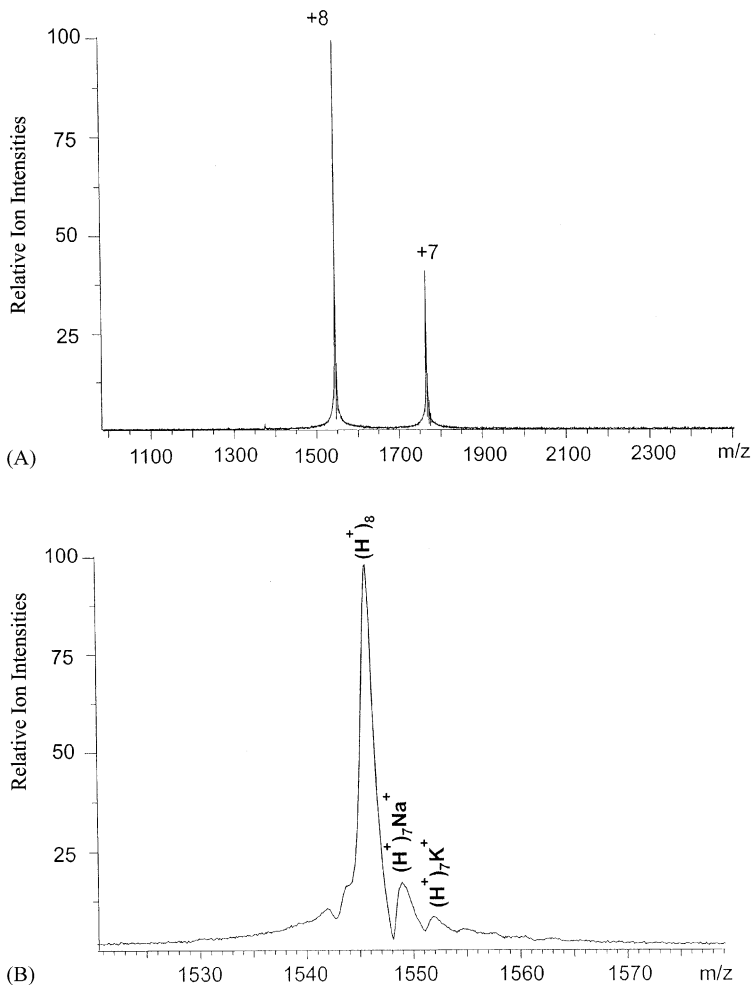


Fig. 10. (A) Mass spectrum observed with CYC = 10^{-5} M, NH_4Ac 10^{-3} M aqueous solution, pH = 6.8. (B) Expanded mass range of $Z = 8$ shows that this is due almost completely to the $(H^+)_8$ ion.

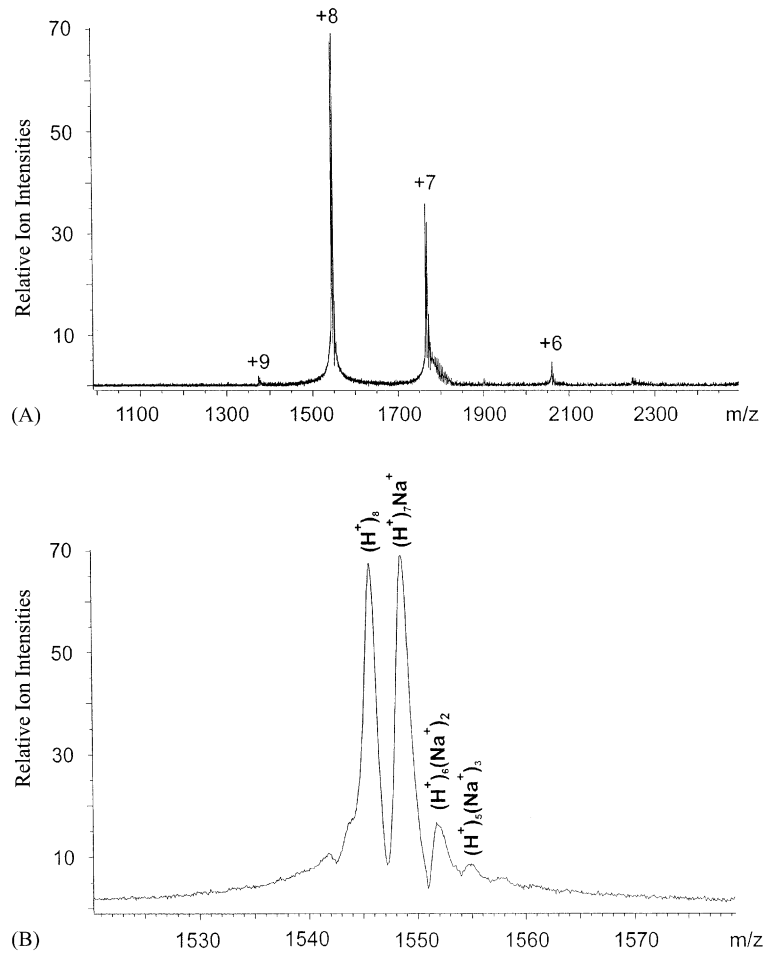


Fig. 11. (A) Mass spectrum observed with CYC 10^{-5} M, NH_4Ac 0.5×10^{-3} M, NaCl 0.5×10^{-3} M aqueous solution, pH = 6.8. (B) Expanded mass spectrum of $Z = 8$ charge state shows the presence of sodiation due to NaCl present in the solution. Surprisingly, observed sodiation is much lower than was the case for CAII, see Fig. 7B.

concentrations, see Fig. 7B. The pure H_{11} peak for CAII is a very minor one, while for CYC, Fig. 10B, the pure H_8 peak is dominant. The difference observed seemed so surprising that repeated measurements with newly prepared solutions were made, which confirmed the results in Fig. 11. The finding that CYC takes up much less Na^+ than CAII shows that the low intensity of Na^+ and K^+ adducts in Fig. 9 is not necessarily due to lower salt contamination of the commercial sample but to structural differences. The absence of α-helix and β-sheets which is a distinguishing feature of CYC, could be responsible for the low M^+ uptake. This observation deserves further study.

The results (Figs. 7–11) do demonstrate clearly, that the observed degree of protonation depends on the presence of NH_4^+, provided by NH_4Ac. When NH_4^+ is present in the solution, protonation is observed, when NH_4^+ is not present, cationization dominates. We consider these observations as a convincing proof that NH_4^+ is the protonating agent of the proteins observed with ESI that use an NH_4Ac buffer to maintain a pH ≈ 7.

3.2. Expected charged states on protonation by NH_4^+

3.2.1. General considerations

According to the CRM proposed for globular proteins [11], the number of charges, Z_{CRM}, on the precursor droplets corresponds to the number of charges that will be transferred to the globular protein. However, for these charges to be observed in the protein mass spectrum there is also the condition that the protein should have a sufficient number of strong basic side chains at its surface that can accept and hold the charges in the "desolvation" process of heating and CAD. We will call the number of such strongly basic side chains N_{SB}.

The fact that the mass spectrometrically observed number of charges Z_{obs} for most proteins obey the relationship in Fig. 2, shows that

$$Z_{CRM} = Z_{obs} \tag{10a}$$

for most proteins.

Now it is extremely unlikely that all the proteins that obey Eq. (10a) have exactly the same number of strong basis sites as Z_{CRM}. It is much more likely that the majority of these proteins have more strong basic sites than required, i.e.,

$$Z_{obs} \approx Z_{CRM} \leq N_{SB} \tag{10b}$$

Eq. (10b) provides the general condition for Eq. (10a) to be obeyed.

However, exceptions where $N_{SB} < Z_{CRM}$ can be expected for globular proteins, which for some special biological purpose have a very small number of basic side chains. In that special case:

$$Z_{obs} < Z_{CRM}$$

such that

$$Z_{obs} \approx N_{SB} \tag{11}$$

It should be noted that the number of "strong basic sites" N_{SB} for multiply-charged proteins depends not only on the base strength of the basic side chains, but also on the number of charges (protons) present on the protein and on the distances between the protonated bases, because the presence of nearby charges facilitates, by Coulombic repulsion, the deprotonation of a given basic site.

The charging via NH_4^+ was discussed in Section 1, see Eq. (7). Only a partial proton transfer to the basic groups occurs and the NH_3 molecule remains attached by a strong hydrogen bond. On "desolvation," the NH_3 molecule will be detached, Eq. (8a). However, charge loss can also occur by deprotonation, i.e., loss of NH_4^+, Eq. (8b). The charge loss (8b) will be promoted by the Coulombic repulsion of nearby charges and can occur even though the gas phase basicity of NH_3 is lower than that of the basic side chains such as Lys.

Williams and coworkers [24] were the first to calculate N_{SB} values. Their work was focused on denatured proteins but results for one native protein, CYC (see Tables 2–4) [24d] were published also. It was the pioneering work of Williams that encouraged us towards calculations of N_{SB} values. The results for CYC were evaluated [24d] for H_3O^+ and not NH_4^+ as the protonating agent.

The procedure used by Williams and coworkers [24] developed first an equation for a simple two proton sites case. This equation was then expanded to cover multiple protonation sites as occur in proteins. For the two proton case, the diprotonated alkyldiamines $H_3N(CH_2)_nNH_3^{2+}$ were used. They may be considered as a simple model for two protonated lysines on opposite sides of a small protein.

The equation proposed [24] for the diprotonated system has been superseded by more recent work due to Gronert [25] which has provided an insightful theoretical analysis of diprotonated alkyl diamines. The potential energy diagram for the diprotonated diamine with $n = 7$ which has one NH_3 adduct, obtained from ab initio calculations [25], is shown in Fig. 12A. For the present discussion we can assume that the ammoniated and diprotonated compound was obtained by ESI and is now heated to achieve "desolvation" which involves the dissociation of the last NH_3 molecule. The activation energy, E_{DP}, for the deprotonating exit

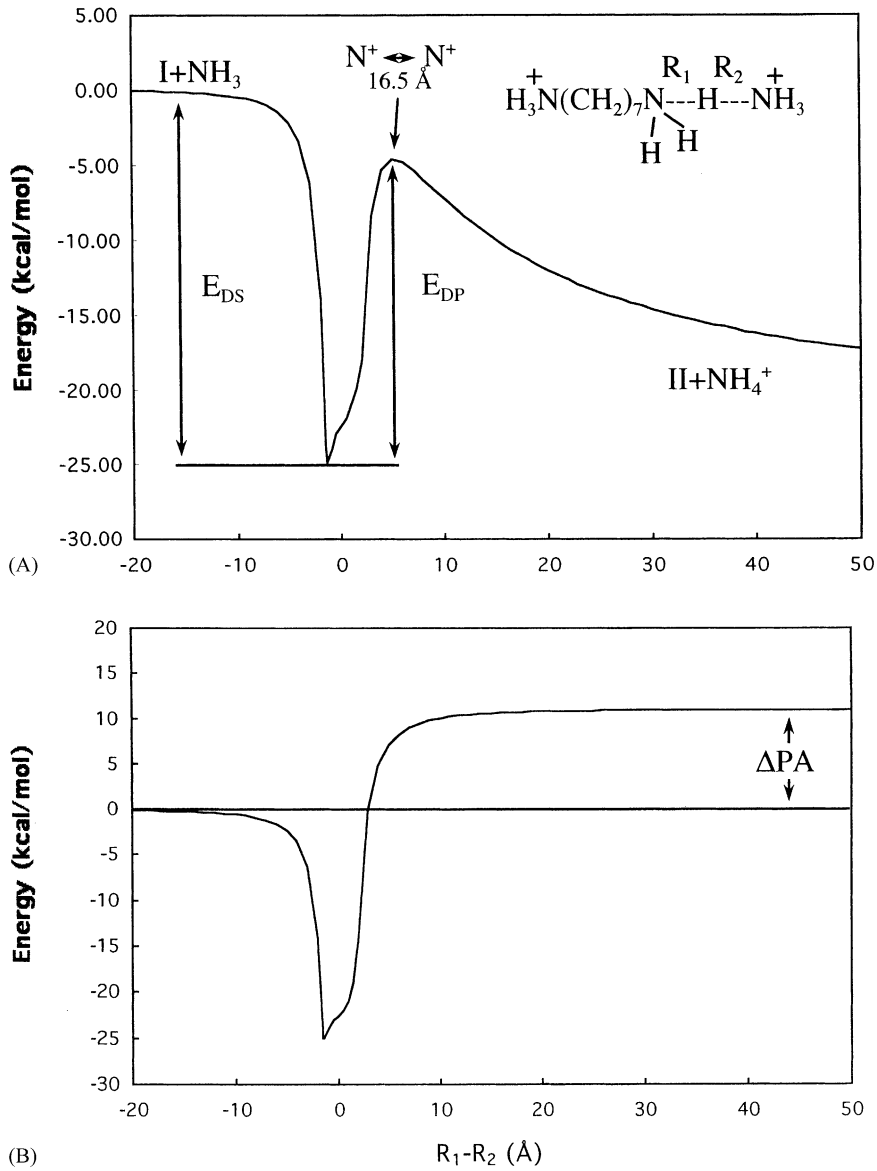


Fig. 12. (A) Energy surface diagram for reaction $NH_3 + (H_3N(CH_2)_7NH_3)^{2+} \rightarrow (NH_3(CH_2)_7NH_3NH_3)^{2+} \rightarrow (NH_3(CH_2)_7NH_2)^+ + NH_4^+$. Evaluated from ab initio calculations by Gronert [25]. The N–N distance of 16.5 Å corresponds to the distance between the charged nitrogen of the diamine and the nitrogen on the NH_3 molecule in the transition state. The distance between the two charged nitrogens of the reactant $NH_3(CH_2)_7NH_3^{2+}$, is $r = 10$ Å. (B) Energy surface diagram for reaction involving singly charged ion $NH_3 + H_2N(CH_2)_7NH_3^+ \rightarrow (NH_2(CH_2)_7NH_3NH_3)^+ \rightarrow NH_2(CH_2)_7NH_2 + NH_4^+$, after Gronert [25].

to the right (Fig. 12A), is lower than the desolvation energy, E_{DS}, and therefore on heating, one expects deprotonation to dominate. The deprotonation activation energy is lower only because of the assistance of the Coulombic repulsion between the two charges.

The reaction coordinate, in the absence of a second charge, is well known. Transition from the left to right

side of the coordinate corresponds to proton transfer from one base to the other

$$BH^+ + B_S = B + B_SH^+$$

For the special case of interest here, the proton transfer is:

$$H_2N(CH_2)_7NH_3^+ + NH_3 = H_2N(CH_2)_7NH_2 + NH_4^+ \quad (12)$$

An ab initio calculated [25] surface for such a reaction is shown in Fig. 12B. Note that E_{DP} is higher than E_{DS}, which means that desolvation (loss of NH_3), and not deprotonation will occur in the absence of Coulombic repulsion by the second charge. The difference, $E_{DP} - E_{DS}$, is positive and corresponds to the proton affinity (PA) difference:

$$E_{DP} - E_{DS} = PA(B) - PA(B_S) \quad (13)$$

where the proton affinity is defined by the energy change ΔE for the reaction:

$$BH^+ = B + H, \quad \Delta E = PA(B) \quad (14)$$

Proton affinities and gas phase basicities, GB, which correspond to the free energy change for Reaction (14), are available [26], for very many compounds including amino acids and compounds that model the side chains, such as *n*-$BuNH_2$ which models the Lys side chain, see Table 1.

The results in Fig. 12A provide the answer, that the number of strong basic sites for $NH_2(CH_2)_7NH_2$, when NH_4^+ is the protonating reagent, is $N_{SB} = 1$. For the $(CH_2)_{12}$ diamine $N_{SB} = 2$, because the distance between the two charges is bigger. In this case, heating leads predominantly to loss of NH_3 and not NH_4^+. For details see Peschke et al. [27].

To obtain N_{SB} values for the proteins of interest, it is desirable to generalize the two charge result (Fig. 12A) to distances other than the N–N distance of 10 Å which is present for the $(CH_2)_7$ diamine [25]. Also, bases modeling peptide side chains, other than Lys, should be involved as well as "solvents" other than NH_3, such as H_2O or MeOH.

Gronert [25] showed that one can do this, on the basis of a reaction coordinate for the singly charged proton transfer (Fig. 12B). This coordinate can be converted to the coordinate for the system with two charges analogous to that in Fig. 12A, by introducing Coulomb equation terms which account for the electrostatic repulsions caused by the second charge. It was shown [25] that such an approach leads to a reaction coordinate in agreement with the accurate, theoretically calculated one (Fig. 12A).

However, the complete reaction coordinate for the two charge case is not really required. The activation energy difference:

$$\begin{array}{ll} E_{DP} - E_{DS}, & \text{two charges} \\ E_{DP} - E_{DS} > 0, & \text{desolvation} \\ E_{DP} - E_{DS} < 0, & \text{deprotonation} \end{array} \quad (15)$$

is the only information needed for obtaining the result, whether, $N_{SB} = 2$ or 1, and this can be obtained without the detailed consideration of the reaction coordinates at all values for $R_1 - R_2$ (Fig. 12). Only energies at or near the maximum values of the activation barriers are required.

This realization [27], led to an examination of the reaction coordinate, Fig. 12A, which revealed that the energy in the region of the activation energies E_{DP} and E_{DS}, can be obtained to a very good approximation

Table 1
Gas phase basicities[a]

	GB_{exp}[b]	$GB_{int,\ protein}$[c]
Arginine	241	251
Histidine	227	237
Imidazole	217	n.a.
Lysine	227	237
n-Butylamine	212	n.a.
Tryptophan	219	234
Indole	216	n.a
Proline	212	227
Ammonia	196	n.a

[a] All values are in kcal/mol.

[b] All values are taken from the NIST database. For those amino acids where side chain GB were available they are shown beneath the amino acid.

[c] Estimated average intrinsic GB of amino acid in protein. The deviation for specific amino acids in the protein is estimated to be on the order of ±5 kcal/mol. Unusual protein arrangements may lead to larger deviations.

entirely from electrostatic calculations [27]. With this approach one can obtain numerical values for $E_{DP} - E_{DS}$, for any distance between the charges and for different bases B_S and B whose proton affinities or gas-phase basicities (Table 1) are known.

For multiply-charged proteins with more than two charges, the two charge pair repulsions can be evaluated and then summed over all charges to obtain a prediction for the maximum number of charges, N_{SB}, that a given protein can hold. These are given in Tables 2–4 for CAII, CYC and pepsin. Additional information concerning the calculations leading to the results presented in Tables 2–4 can be found in Peschke et al. [27].

3.2.2. Carbonic anhydrase, Table 2

The major charge state is $Z_{obs} = 11$ (Fig. 4). The density of protein, $\rho = 1\,g/cm^3$, assumed by Fernandez de la Mora leads to a radius, $R = 22.8$ Å [11]. A similar radius $R = 23$–25 Å can be deduced by inspection of the X-ray structure [21]. The

Table 2
Comparison of observed charges, Z_{obs}, with, N_{SB}, the calculated number of protons that protein can hold and the charges provided by CRM (carbonic anhydrase, MW, 29,200)[a,b]

	Charges, Z	Ammonia, $E_{DP} - E_{DS}$	Water, $E_{DP} - E_{DS}$	Residue[c,d]
Z_{obs}[e]	11	9	47	Lys133
Z_{CRM}[f]	12	8	46	Lys113
	13	6	44	Lys168
	14	3	41	Lys213
N_{SB}[g]	15	1	39	Lys252
	16	−2	36	Lys170
	17	−4	34	Lys159
	18	−5	33	Lys261
	19	−8	30	Lys80
	20	−13	25	Lys127

[a] All energies are given in kcal/mol.
[b] Total of 36 basic sites considered.
[c] Residue that is easiest to be deprotonated in presence of a base such as NH_3 or H_2O.
[d] Protonated bases for the first nine charges (in addition to $Zn(OH)^+$): His15, Arg27, His36, Arg58, Arg89, Lys154, Arg182, Arg227, Lys257.
[e] Observed charge with highest intensity (see Fig. 4).
[f] Charge provided by CRM from precursor droplet.
[g] Largest number of charges that available basic sites can hold.

Table 3
Comparison of observed charges, Z_{obs}, with, N_{SB}, the calculated number of protons that protein can hold and the charges provided by CRM (cytochrome *c*, MW 12,400)[a,b]

	Charges, Z	Ammonia, $E_{DP} - E_{DS}$	Water, $E_{DP} - E_{DS}$	Residue[c,d]
Z_{obs}[e]	8	10	48	Lys53
Z_{CRM}[f]	9	5	43	Lys27
N_{SB}[g]	10	1	39	Lys99
	11	−1	37	Lys22
	12	−3	35	Lys87
	13	−9	29	Lys79
	14	−15	23	Lys60
	15	−17	21	Lys8
	16	−21	17	Lys39
	17	−26	12	Lys100

[a] All energies are given in kcal/mol.
[b] Total of 29 basic sites considered.
[c] Residue that is easiest to be deprotonated in presence of a base such as NH_3 or H_2O.
[d] Protonated bases for the first seven charges: Lys7, Lys13, His26, Arg38, Lys55, Lys72, Arg91.
[e] Observed charge with highest intensity, see Fig. 10.
[f] Charge provided by CRM from precursor droplet.
[g] Largest number of charges that available basic sites can hold.

Table 4
Comparison of observed charges, Z_{obs}, with, N_{SB}, the calculated number of protons that protein can hold and the charges provided by CRM (pepsin, MW 34,600)[a,b]

	Charges, Z	Ammonia, $E_{DP} - E_{DS}$	Water, $E_{DP} - E_{DS}$	Residue[c,d]
	8	3	41	Pro292
N_{SB}[e]	9	1	39	Pro116
Z_{obs}[f]	10	−1	37	Pro23
	11	−3	35	Pro268
	12	−4	34	Pro256
	13	−11	27	Pro183
	14	−16	22	Pro135
Z_{CRM}[g]	15	−19	19	Pro324
	16	−23	15	Trp141
	17	−33	5	His53

[a] All energies are given in kcal/mol.
[b] Total of 24 basic sites considered.
[c] Residue that is easiest to be deprotonated in presence of a base such as NH_3 or H_2O.
[d] Protonated bases for the first seven charges: Pro5, Pro58, Pro108, Trp181, Trp299, Arg308, Arg315.
[e] Largest number of charges that available basic sites can hold.
[f] Standing and coworkers [22]. Pepsin itself is difficult to observe with ESI and values of $Z_{obs} = 10$ are an average of reported data [22].
[g] Charge provided by CRM from precursor droplet.

Rayleigh equation leads to a $Z_R \approx 13.6$ for this radius, however as already pointed out by Fernandez de la Mora [11], water droplets with radius lower than $R \approx 80$ Å have a charge at their surface which is somewhat lower than the Rayleigh charge, because such small droplets loose charge by the ion evaporation model (IEM). The charge on droplets losing small ions by IEM, cannot be evaluated accurately, but is some 10–20% lower than the charge expected under Rayleigh fission conditions [8,10,11,17]. $Z_{CRM} = 12$ as given in Table 2 is thus a "best" choice.

The evaluated number of charges, $N_{SB} = 15$, indicates that the protein can hold many more charges than the $Z_{CRM} = 12$ charges, that need to be accommodated. If H_2O rather than NH_3 was involved, an $N_{SB} \gg 20$, see Table 2, is predicted. Thus, the data obtained for CAII are consistent with the assumption that in general N_{SB} is higher than Z_{CRM} so that Z_{CRM} is the charge limiting factor (see Eq. (9)).

3.2.3. Cytochrome c, Table 3

The most intense charge state is $Z_{obs} = 8$ (Fig. 10). The Rayleigh charge for a protein density $\rho = 1$ g/cm^3 is $Z_R = 9$. The X-ray structure shows that CYC is not spherical, which indicates that a larger charge such as $Z_R = 10$, should be assigned. However, the surface charge in this smaller protein is expected to be smaller since it is controlled by IEM. Best choice $Z_{CRM} = 9$. The predicted maximum charge that can be held when NH_3 is involved is $N_{SB} = 10$, if water was involved $N_{SB} > 17$. These CYC data are also consistent with the proposition that for most proteins Z_{CRM} is the charge determining factor (see Eq. (9)).

3.2.4. Pepsin, Table 4

Pepsin was deliberately chosen. It is a digestive enzyme that is designed to function in an acidic medium and has an extremely large number (~30) acidic (Glu and Asp) side chains near its surface. On the other hand it has only four strongly basic side chains, Lys319, His53, Arg315, and Arg307, at its surface. Therefore, this special protein is expected to be an exception of the rule that Z_{CRM} determines Z_{obs} (because N_{SB} is much higher, Eq. (9)). For pepsin one expects $Z_{obs} \approx N_{SB}$, because the number of basic sites are the limiting factor, Eq. (11).

Unfortunately, pepsin is difficult to produce by ESI–MS at pH 7. We were not able to observe useful mass spectra under these conditions. The $Z_{obs} = 10$ is due to results by Standing and coworkers [22]. These authors also had difficulties observing pepsin, but were able to observe pepsin from ESI of the complex of pepsin and its inhibitor pepstatin. These indicate a $Z_{obs} = 10$ for pepsin, which is the value used in Table 4. The shape of pepsin is also significantly different from that of a sphere, which makes the evaluation of Z_{CRM} ambiguous. A value of $Z_R \approx 16$, can be obtained from assuming density of 1 g/cm^3. A value of $Z_R \approx 16$ can be obtained also by selecting a radius that would fit the shape of pepsin observed in the X-ray structure. This value should be reduced by some 10% because the charge is controlled by IEM. Thus, the "best" choice of $Z_{CRM} = 15$ is used in Table 4.

The result that the evaluated $N_{SB} = 9$ and $Z_{obs} \approx 10$, provides very good evidence that for the special case where basic groups are scarce, the charge limiting factor is N_{SB}.

3.3. Cleanup in the desolvation stage

The formation of globular proteins by CRM is expected to lead not only to a given multiple charging of the protein, but also to deposition of other, neutral solutes on the protein. It can be demonstrated that rather large quantities of solutes could be deposited.

Consider a litter of droplets that have just formed by a Rayleigh fission of a precursor droplet (Fig. 1). One or more of these droplets contain one globular protein. Assume that the droplets that were just formed have a charge Z and a radius R_i. The charge Z is too small to lead to an immediate new fission. For the average droplet it is approximately 0.7 times too small [11]. The droplets shrink, by evaporation of solvent, from the initial radius R_i, to the radius R_R when a Rayleigh

fission can occur. Eq. (16) can be obtained:

$$\frac{R_i}{R_R} = \left(\frac{1}{0.7}\right)^{2/3} = 1.27, \qquad \frac{V_i}{V_R} = \left(\frac{R_i}{R_R}\right)^3 \approx 2 \tag{16}$$

with use at the Rayleigh Eq. (2) and the condition that the charge is constant. The result is that the volume of the initial droplet will be approximately twice as large as the volume at the Rayleigh limit. The volume of the Rayleigh limit should approximately correspond to the volume of the protein, because of the relationship observed in Fig. 2.

A common choice of concentrations in the solution to be sprayed is $\sim 10^{-5}$ M of the protein and $\sim 10^{-3}$ M or more of the buffer. The ratio 100:1 moles buffer to moles protein is expected to be approximately preserved in the droplet under consideration. According to Eq. (16) we can expect some 100 molecules of ammonium acetate (NH_4Ac) to land on the protein on evaporation of the precursor water droplet. The desolvation stage would have to be able to remove such a deposit. Evidently this is possible with NH_4Ac as buffer because use of NH_4Ac leads to observation of clean protein mass spectra.

That a rather large number of solute molecules originating from the buffer can be deposited is illustrated by the experiments of Smith and coworkers [2b], who observed carbonic anhydrase molecule spectra (charge state 10) which had up to 40 citric acid molecule adducts when the buffer used was ammonium citrate. The exact origin of the neutralized citric acid adduct observed [2b] cannot be easily predicted, but possibly, on evaporation of the water and during the desolvation process, loss of ammonia (NH_3) could have occurred which left behind the very much less volatile citric acid. Acetic acid, on the other hand, is much more volatile and on heating of ammonium acetate, one might expect that both NH_3 and acetic acid are formed, which easily evaporate.

In the present experiments where sodium chloride and other MX salts were added (see Figs. 6, 7 and 9) at concentrations up to 0.5 mM MX, deposits of some 10–50 or more MX molecules on one protein molecule might be expected. But the presence of only a few MX adducts was observed in the spectra. Thus, the cleanup of MX during the desolvation stage must have been quite efficient, yet one might have expected that these ion-paired species are not sufficiently volatile.

Shown in Table 5 are the binding energies [27,28–31] for several model compounds which are relevant. The binding enthalpies of H_2O and NH_3 to the diprotonated diamines illustrate the bond strength of these molecules to protonated sites, which model protein sites. The bond energy values were obtained from ion–molecule equilibria measurements in the gas phase [26,28,30] which allow also the observation of the temperature range over which the dissociation of the solvent molecule is fast (tens of microseconds). These temperatures are also given in Table 5. Removal of H_2O and NH_3 thus requires temperatures in the 100–125 °C range. The binding energies for these adducts are in the 20 kcal/mol range and these adducts are easily removed.

Na^+ and NaCl are expected to interact with the peptide oxygens. For these interactions N–Me acetamide ($CH_3CONHCH_3$) is a good model. The Na^+ bond to the oxygen of N–Me acetamide is 36 kcal/mol. This is a very strong bond and Na^+ is not expected to be lost during the dissolvation stage. Furthermore, the Na^+ ion might even be more strongly bonded by dicoordination to two peptide oxygens of the protein. Yet some loss of M^+ ions on cleanup was indicated, see Fig. 9, which shows a small shift to lower charge states as protonation is replaced by cationization by M^+.

Loss of Na^+ (or M^+ in general) is however possible through the mediation by MCl. Thus, some NaCl may deposit on Na^+ coordinated to peptide oxygens. The results in Table 5, which give the bond energies of Na^+–acetone, 31 kcal/mol and $NaClNa^+$–acetone, ≈23 kcal/mol, illustrate the effect and the likelihood that the loss of M^+ can be mediated by MX, just like the loss of a proton was mediated by NH_3.

A decrease in the charge state when M in MX was changed from Na to Cs was observed in Fig. 9. However, this decrease was not very big. Had only the bonding of M^+ to the peptide oxygen been involved, one could have expected a larger decrease of charge state, because bonding to the larger Cs^+ is very much

Table 5
Bond energies of model compounds relevant to cleanup of proteins

Reaction	Energy (kcal/mol)	Temperature (°C)
$(NH_3(CH_2)_8NH_3–OH_2)^{2+} = NH_3(CH_2)_8NH_3^{2+} + OH_2$	16.9[a]	∼110[b]
$(NH_3(CH_2)_{12}NH_3–OH_2)^{2+} = NH_3(CH_2)_{12}NH_3^{2+} + OH_2$	15.7[a]	∼110[b]
$(NH_3(CH_2)_{12}NH_3–NH_3)^{2+} = NH_3(CH_2)_2NH_3^{2+} + NH_3$	19.5[c]	∼125[b]
$(CH_3)_2CO–Na^+ = (CH_3)_2CO + Na^+$	31.2[d], 34.3[e]	300[f]
$(CH_3)_2CO–NaCl = (CH_3)_2CO + NaCl$	17.8[e]	120[b]
$(CH_3)_2CONaClNa^+ = (CH_3)_2CO + NaClNa^+$	26.5[e]	150
$NaClNa^+ = NaCl + Na^+$	47.7[e]	>300[f]
$(CH_3)_2COK^+ = (CH_3)_2CO + K^+$	26.0[g]	∼250[f]
$CH_3CONHCH_3K^+ = CH_3CONHCH_3 + K^+$	30[g]	≈350[f]
$CH_3CONHCH_3Na^+ = CH_3CONHCH_3 + Na^+$	36[h]	>400[f]

[a] From equlibria [28].
[b] Fast dissociation (half-life 10–100 μs) observed from equilibria of reaction shown at nitrogen gas pressures 10 Torr. Expected to dissociate at cleanup temperatures below 180 °C.
[c] From equilibria [28].
[d] Armentrout and Rodgers [29].
[e] Calculated ab initio values, present work, B3LYP/6-311++G(d,p) level without zero point energy correction. Corrected values expected to be some 3 kcal/mol lower.
[f] Decomposition not expected to be fast and lead to removal at cleanup temperature below 250 °C.
[g] Equilibrium values [30].
[h] Experimental and theoretical bond energy values [31].

weaker, i.e.,

>C=O · · · M^+ bond energy decreases rapidly from Na^+ → Cs^+

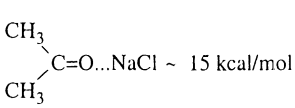
However, when MXM^+ is present,

>C=O · · · MXM^+ bond energy decreases slowly from $Na^+ \rightarrow Cs^+$

because a small M^+, bonds strongly also to MX and bonding to MX facilitates the decomposition.

The observed, very low uptake, of Na^+ for CYC, Fig. 11, compared to that for CAII, Fig. 7, may be due to the fact that CYC has no α-helices and β-chains and thus may be less suited to provide dicoordinated sites for the Na^+ ion, which in general are expected to foster retention of the Na^+ on protein.

Neutral NaCl molecules, which may not find a place on Na^+ coordinated sites, may settle on peptide oxygens. The bond energy, see Table 5,with zero point correction, is very small and the bonding to *N*-methyl acetamide, which models the peptide although stronger is still expected to be under 20 kcal/mol so that temperatures of 100–150 °C in the desolvation stage should easily remove the NaCl adducts.



While the data are qualitative, they do illustrate that a rational account of the possible fate of various adducts resulting from CRM can be given.

4. Conclusions

(1) Native globular proteins sprayed at pH ≈ 7, are produced by the CRM. The charges at the surface of the evaporating droplet that contains one protein molecule are due to small ions originally present in the solution used. Charging of the protein occurs via these surface ions.
(2) When the predominant electrolyte leading to small ions is the buffer, such as ammonium acetate (NH_4Ac), it is the excess NH_4^+ ions at the surface of the droplet that are involved in the protonation of the protein. This can be

demonstrated by using mixtures of NH_4Ac and NaCl or other alkali ion salts (MX). When the concentration ratio NH_4Ac to MX is small, little protonation is observed and protonation is replaced by cationization by M^+.

(3) NH_4^+ can be the protonating agent because the gas phase basicity of NH_3 is lower than the gas phase basicities of side chains such as Arg, Lys, His, Trp, Pro.

(4) Fernandez de la Mora's proposal that

$$Z_{obs} \approx Z_{CRM}$$

where Z_{obs} are the mass spectrometrically observed charges on a globular protein, and Z_{CRM} are the number of excess charges expected to be at the surface of a water droplet, with the same radius as the protein (Fig. 2), is supported by evaluation of N_{SB}, the number of basic sites at the surface of the protein which have sufficient basicity to hold a maximum charge: $Z_{max} = N_{SB}$. The proteins, carbonic anhydrase (CAII) and CYC, had a $N_{SB} > Z_{obs}$. This shows that these proteins can accommodate all the charges that the droplet provides such that, $Z_{obs} \approx Z_{CRM}$.

(5) One protein, pepsin, is an exception to the relationship $Z_{CRM} \approx Z_{obs}$, and has a Z_{obs} that is significantly lower than Z_{CRM}. It is shown that pepsin has an insufficient number of basic sites, N_{SB}. For this protein, $N_{SB} \approx Z_{obs}$. Proteins with as few basic sites at the surface as pepsin are the exception, rather than the rule.

(6) The formation of the protein in the gas phase via CRM leads not only to charging but also to deposition of neutral adducts due to the presence of solutes such as buffers in the precursor droplet. Estimates based on CRM predict values for the number of solute molecules expected to be deposited on the protein.

(7) Bond energies of model solute–protein interactions can provide estimates as to which solutes are sufficiently volatile so that they can be removed by the desolvation-cleanup process achieved by heating and CAD.

(8) The formation of charged proteins in the gas phase via CRM is very gentle and is expected to lead to relatively small changes of structure in the transition to the gas phase. The desolvation-cleanup stage may be a source of changes. Careful choice of buffer and purity of solution and protein sample may be used to minimize the need for a high temperature cleanup stage.

Acknowledgements

The FT-ICR used was an instrument in Professor J. Klassen's laboratory, who generously allowed us to use it. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] (a) B. Ganem, Y. Li, J.D. Henion, J. Am. Chem. Soc. 113 (1991) 6294;
(b) B. Ganem, Y.T. Li, J. Henion, J. Am. Chem. Soc. 113 (1991) 7818;
(c) V. Katta, B.T. Chait, J. Am. Chem. Soc. 113 (1991) 8534;
(d) M. Baca, S.B.H. Kent, J. Am. Chem. Soc. 114 (1992) 3992;
(e) A.K. Ganguly, B.N. Pramnik, A. Tsarbopoulos, T.R. Covey, E. Huang, S.A. Fuhrman, J. Am. Chem. Soc. 114 (1992) 6559;
(f) K.J. Light-Wahl, D.L. Springer, B.E. Winger, C.G. Edmonds, D.G. Camp II, B.D. Thrall, R.D. Smith, J. Am. Chem. Soc. 115 (1993) 803;
(g) B. Ganem, Y.T. Li, J.D. Henion, Tetrahedron Lett. 34 (1993) 1445;
(h) R.R. Ogorzlek Loo, D.R. Goodlett, R.D. Smith, J.A. Loo, J. Am. Chem. Soc. 415 (1993) 4391;
(i) M. Jacquinod, E. Leize, N. Potier, A.M. Albrecht, A. Shanzer, A. Van Dorsselaer, Tetrahedron Lett. 34 (1993) 2771;
(j) E.C. Huang, B.N. Pramanik, A. Tsarbopoulos, P. Reichert, A.K. Ganguly, P.P. Trotta, T.L. Nagabhashan, T.R. Covey, J. Am. Soc. Mass Spectrom. 4 (1993) 624;
(k) Y.T. Li, Y.L. Hsieh, J.D. Henion, B. Ganem, M.W. Senko, F.W. McLafferty, J. Am. Chem. Soc. 115 (1993) 8409;
(l) J.A. Loo, R.R. Ogorzalek Loo, P.C. Andrews, Org. Mass Spectrom. 28 (1993) 1640;
(m) B.S. Schwartz, K.J. Light-Wahl, R.D. Smith, J. Am. Soc. Mass Spectrom. 5 (1994) 201;
(n) K.J. Light-Wahl, B.L. Schwartz, R.D. Smith, J. Am. Chem.

Soc. 116 (1994) 5277;
(o) D. Lafitte, Eur. J. Biochem. 261 (1999) 337.
[2] (a) R.D. Smith, J.A. Loo, R.R. Ogorzalek, M. Busman, H.R. Udset, Mass Spectrom. Rev. 10 (1991) 359;
(b) L.P. Tolic, J.P. Bruce, Q.P. Lei, G.A. Anderson, R.D. Smith, Anal. Chem. 70 (1998) 405;
(c) J.E. Brüce, S.L. Van Orden, G.A. Anderson, S.A. Hofstadler, M.G. Sherman, A.L. Rockwood, R.D. Smith, J. Mass Spectrom. 30 (1995) 124;
(d) R.D. Smith, K. Tang, H. Udseth, A.L. Rockwood, in: Proceedings of the 44th American Society Mass Spectrometry Conference, 1996, p. 1086.
[3] L. Rayleigh, Philos. Mag. 14 (1882) 184.
[4] A. Gomez, K. Tang, Phys. Fluids 6 (1994) 404.
[5] I. Hayati, A.I. Bailey, T.F. Tadros, J. Colloid Interf. Sci. 117 (1987) 205.
[6] J. Fernandez de la Mora, I.G. Locertales, J. Fluid Mech. 243 (1994) 561.
[7] (a) P. Kebarle, L. Tang, Anal. Chem. 65 (1993) 972A;
(b) L. Tang, P. Kebarle, Anal. Chem. 65 (1993) 3654.
[8] P. Kebarle, M. Peschke, Anal. Chim. Acta 406 (2000) 11.
[9] G. Wang, R.B. Cole, Anal. Chim. Acta 406 (2000) 53.
[10] M. Gomero-Castano, J. Fernandez de la Mora, Anal. Chim. Acta 406 (2000) 67.
[11] J. Fernandez de la Mora, Anal. Chim. Acta 406 (2000) 93.
[12] M. Labowsky, J.B. Fenn, J. Fernandez de la Mora, Anal. Chim. Acta 406 (2000) 105.
[13] R.B. Cole, J. Mass Spectrom. 35 (2000) 773.
[14] M. Gamero-Castano, J. Fernandez de la Mora, J. Mass Spectrom. 35 (2000) 780.
[15] P. Kebarle, J. Mass Spectrom. 35 (2000) 804.
[16] (a) M. Dole, L.L. Mack, R.L. Hines, R.C. Molbley, L.D. Ferguson, M.B. Alice, J. Chem. Phys. 49 (1968) 2240;
(b) L.L. Mack, P. Kralic, A. Rehude, M. Dole, J. Chem. Phys. 52 (1970) 4977;
(c) J. Glenic, L.L. Mack, K. Nakamal, C. Gupta, V. Cumar, M. Dole, Biomed. Mass Spectrom. 11 (1984) 259.
[17] I.V. Iribarne, B.A. Thomson, J. Chem. Phys. 64 (1976) 2287.
[18] P. Kebarle, Y. Ho, in: R.B. Cole (Ed.), Electrospray Ionization Mass Spectrometry, Wiley, New York, 1977, p. 55.
[19] (a) R.P. Tolic, G.A. Anderson, R.D. Smith, H.M. Brothers, R. Spindler, D.A. Tomalia, Int. J. Mass Spectrom. Ion Proc. 165 (1997) 405;
(b) I.V. Chernishevich, W. Ens, K.G. Standing, in: W. Ens, K.G. Standing, I.V. Chernushevich (Eds.), New Methods for the Study of Biomolecular Complexes, Kluwer Academic Publishers, Dodrecht, Boston, London, 1998, p. 101.
[20] (a) J.C.Y. Le Blanc, J. Wang, R. Guevremont, K.W.M. Siu, Org. Mass Spectrom. 29 (1994) 587;
(b) A.T. Ivarone, J.C. Jurchen, E.R. Williams, J. Am. Soc. Mass Spectrom. 11 (2000) 976.
[21] P.O. Appel, A. Bairoch, D.F. Hochstrasser, ExPASy Database, Swiss Bioinformatics, Trends Biochem. Sci. 19 (1994) 258.
[22] I.V. Chernushevich, W. Ens, K. Standing, in: R.B. Cole (Ed.), Electrospray Ionization Mass Spectrometry, Wiley, New York, 1997, p. 223.
[23] T. Takano, O.B. Kallai, R. Swanson, R.E. Dickerson, J. Biol. Chem. 248 (1973) 5244.
[24] (a) D.S. Gross, S.E. Rodriguez-Cruz, E.R. Williams, J. Phys. Chem. (1995) 4034;
(b) D.S. Gross, E.R. Williams, J. Am. Chem. Soc. 117 (1995) 883;
(c) P.D. Schnier, D.S. Gross, E.R. Williams, J. Am. Chem. Soc. 117 (1995) 672;
(d) P.D. Schnier, D.S. Gross, E.R. Williams, J. Am. Soc. Mass Spectrom. 6 (1995) 1086;
(e) E.R. Williams, J. Mass Spectrom. 31 (1996) 831.
[25] (a) S. Gronert, J. Am. Chem. Soc. 118 (1998) 3525;
(b) S. Gronert, J. Mass Spectrom. 34 (1999) 787.
[26] E.P. Hunter, S.G. Lias, J. Phys. Chem. Ref. Data 27 (1998) 3.
[27] M. Peschke, A.T. Blades, P. Kebarle, J. Am. Chem. Soc., submitted for publication.
[28] J.S. Klassen, A.T. Blades, P. Kebarle, J. Am. Chem. Soc. 99 (1995) 15509.
[29] P.B. Armentrout, M. Rodgers, J. Phys. Chem. A 104 (2000) 2238.
[30] J. Sunner, P. Kebarle, J. Am. Chem. Soc. 106 (1984) 6135.
[31] J.S. Klassen, P. Kebarle, J. Phys. Chem. 100 (1996) 14218.