Protein–ligand and protein–protein interactions studied by electrospray ionization and mass spectrometry

W.I. Burkitt*, P.J. Derrick*, D. Lafitte† and I. Bronstein‡

*Department of Chemistry, University of Warwick, Coventry CV4 7AL, U.K., †Faculte de Pharmacie, 27 Boulevard Jean Moulin, 13385 Marseille Cedex 05, France, and ‡Institute of Animal Health, Protein Structure Group, Compton Laboratory, Compton, Newbury RG20 7NN, U.K.

Abstract
Electrospray ionization has made possible the transference of non-covalently bound complexes from solution phase to high vacuum. In the process, a complex acquires a net charge and becomes amenable to measurement by MS. FTICR (Fourier-transform ion cyclotron resonance) MS allows these ions to be measured with sufficiently high resolution for the isotopomers of complexes of small proteins to be resolved from each other (true for complexes up to about 100 kDa for the most powerful FTICR instruments), which is of crucial significance in the interpretation of spectra. Results are presented for members of the S100 family of proteins, demonstrating how non-covalently bound complexes can be distinguished unambiguously from covalently bound species. Consideration relevant both to determination of binding constants in solution from the gas-phase results and to the elucidation of protein folding and unfolding in solution are discussed. The caveats inherent to the basic approach of using electrospray and MS to characterize protein complexes are weighed and evaluated.

Introduction
Interest in MS among biochemists has risen sharply in recent years, reflecting progress in developing useful methods for applying this mature instrumental technique to biological macromolecules. The overriding challenge in seeking to apply MS to macromolecules has always been how to vaporize samples. Progress in meeting this challenge has been steady for over 30 years, and different methods have been used at different times. The methods currently used most widely are MALDI (matrix-assisted laser-desorption ionization) [1,2] and ESI (electrospray ionization) [3,4]. Before MALDI and ESI, there was fast-atom bombardment [5], and before that plasma desorption [6], with field desorption [7] before that. All of these methods work well for proteins, and there has always been a slant evident in development work in MS towards proteins rather than, say, nucleic acids and carbohydrates.

The growing realization that characterization of genes yields insufficient information for understanding biological processes has brought greater attention to protein complements of cells and tissues, and has been a major factor generating interest in MS.

MALDI and ESI differ significantly in that the former typically affects sublimation of a crystalline solid, whereas the latter concerns the vaporization of a liquid solution. Thus, if the study is to probe by MS the properties of a biomacromolecule that is dependent upon or involves solvent interactions, the method of choice today is ESI. This paper will focus on protein–ligand and protein–protein interactions studied by ESI. In principle, any mass-spectrometric technique can be used to measure gaseous protein ions formed by ESI, but this paper will concentrate on experiments using FTICR (Fourier-transform ion cyclotron resonance) [8]. The reason for this sharp focus is that FTICR affords resolution (in terms of mass) sufficiently high to resolve 13C isotopic components of small proteins (molecular mass ≤ 100 kDa). FTICR offers mass resolutions higher than those of other mass-spectrometric techniques, although there is the caveat that resolution (more precisely resolution in terms of mass-to-charge ratio) is inversely proportional to mass with FTICR. The importance of resolving 13C isotopomers with ESI is that the spacing of adjacent 13C peaks is related to 1/z (where z is the net charge of the ion in units of electronic charge), and knowing z allows the mass m of an ion to be determined from the measured mass-to-charge ratio m/z. If the 13C isotopic peaks are not resolved, the conundrum that m/z is measured and both m and z are unknown must be resolved by some means in order to interpret the experimental results.

ESI
ESI involves nebulizing a liquid typically at atmospheric pressure with the aid of a high electric field. Barely visible droplets are created from protein solutions, and these droplets shrink through evaporation of solvent. The droplets become unstable and eventually isolated protein ions are obtained.
While these events leading to gaseous ion formation are occurring, the droplets are taken into the mass spectrometer’s vacuum system. The particular ESI source employed for the results reported here is shown in Figure 1.

A point to be considered regarding ESI ion sources is that there is typically a supersonic expansion at some stage. In the source shown in Figure 1, this would occur as the fluid emerged from the capillary. There are various consequences of this. One of the most widely recognized is that a biomacromolecule such as a protein could be damaged as a result of molecular collisions, particularly as the ordered flow of the supersonic beam breaks down at the Mach disc [9]. The possibility of dissociation in this way is of particular concern when seeking to measure non-covalently bound complexes such as protein–ligand or protein–protein complexes. The other possibility, which is less widely recognized, is that aggregates might form as a result of supersonic expansion.

Experimental
All measurements were made using a 9.4T FTICR (Bruker Daltonics), which has been described in [10] and is shown in Figure 2. The superconducting magnet on this instrument was passively shielded.

ESI was performed using an external ESI source (Analytica, Branford, CT, U.S.A.). The source conditions were chosen so that dissociation of the non-covalent complexes in the source were minimized. The drying gas used was nitrogen and it was heated to a temperature of 50°C. The solution conditions used for the ESI were 60 µM murine S100A4 (with the terminal methionine residue cleaved) in 5 mM ammonium acetate buffer (pH 5.9).

Results and discussion
An ESI mass spectrum for the S100A4 protein is shown in Figure 3. The five peaks in the region m/z 900–1500 represent species [M + xH]⁺, where M is the mass of the intact S100 molecule and x has values from 13 to 5. The average molecular mass M of S100A4 is 11,590.21 Da. The enlarged picture of the m/z 1055 peak (labelled A in Figure 3) shows that, at higher magnification, there is a cluster of isotopomeric peaks for each peak in the spectrum. Successive isotopomeric peaks represent species differing from each other by one ¹³C atom (i.e. the species represented by the higher m/z peak has one more ¹³C atom and one less ¹²C atom than the lower m/z species). The difference Δm in mass between
successive isotopomeric peaks is thus 1 Da. Therefore, the spacing between successive peaks in terms of \( m/z \) must be \( 1/x \), where \( x \) is the charge on the ion. Knowing \( x \), the mass of the ion is also known.

These peaks at the highest charge states (\( x \) defines the charge state) are attributed to unfolded conformations of the protein [11]. The hypothesis is that unfolded conformations acquire higher charges during ESI, possibly as they make the final move from droplet to isolated ion. The conformation referred to is that of the protein in solution prior to spraying. Thus the hypothesis demands that the solution-phase conformation be retained qualitatively during spraying and break up of droplets, if, that is, the charge acquisition occurs in the final step. Once in the gas phase, the conformation would be influenced by among other things the charge distribution within the gaseous ion, and would not necessarily have any direct relationship to that in solution [12]. That the conformation in the gas phase probably differs in significant respect from that in solution does not vitiate the above discussion of the spectra, but does represent an important caveat for attempts to interpret gas-phase fragmentation patterns.

The peaks at higher \( m/z \) are composite, in the sense of representing more than one species. Certain of the peaks represent both dimers and monomers, whose isotopic peaks fall on top of each other. Consider a monomer ion of formula \((M + x_1 H)^{x_1+}\) and a dimer ion of formula \((2M + 2x_2 H)^{2x_2+}\), where \( x_2 = 2x_1 \). The peaks for these ions will fall at the same \( m/z \) (i.e. the mean positions will be the same). Adjacent isotopic peaks for the dimer ion will be separated by \((1/2x_2) m/z\), whereas those for the monomer ion will be separated by \((1/x_1) m/z\). Thus the isotopic peaks for the dimer are more closely spaced than those for the monomer. Every other isotope peak for the dimer falls on top at a monomer peak. In between these common positions, there are pure dimer peaks. The cluster of peaks labelled D in Figure 3 show this pattern. The envelopes of isotope peaks for the monomer and dimer ions in Figure 3(D) are centred at the same \( m/z \) if the dimer mass is twice that of the monomer. The dimer mass will be twice that of the monomer, if the two monomers composing the dimer are non-covalently bound to each other, i.e. the mass of the dimer (neglecting proton attachment) is indeed 2M.

The cluster of peaks labelled B in Figure 3 represents a monomer overlapping a dimer in which the monomers are covalently bound to each other. The dimer (Figure 3B) has a mass which is 4 Da less than twice the mass of the monomer, as established by the shift to lower \( m/z \) of the centroid of the dimer’s envelope of peaks away from that of the monomer’s envelope. Thus it is established that S100A4
dimerizes in aqueous solution (pH 5.9) to both a covalently and a non-covalently bound species. The covalent species has two covalent links and these are presumably cysteine links. S100A4 contains four cysteine residues and is able to form both inter- and intramolecular bonds.

The cluster of peaks labelled C in Figure 3 represent dimer ions with a charge of 11+. There can be no overlap with monomer ions because the charge number is odd (a monomer with charge 5.5 is impossible). Comparison of the measured masses of these peaks with the calculated masses of the isotopomers of covalent and non-covalent dimers of S100A4 establishes that this cluster of peaks results from a mixture of covalent and non-covalent dimers.

The covalently bound dimer has a higher charge than the non-covalently bound dimer, 12+ for the former and 10+ for the latter. This would imply that the non-covalent dimer has the more folded structure. The fact that the two dimers have different structures might suggest that there is a completely different form of dimerization for the covalent and non-covalent dimers in S100A4.

The cluster of peaks labelled E in Figure 3 represents a mixture of dimers in the 9+ charge state and tetramers in the 18+ charge state, with the peaks deriving from the isotopomers of tetrameric species having a spacing half of that of the dimeric species. The centroid of the isotopic envelope of the tetrameric peaks is shifted to lower m/z compared with that of the centroid of the isotopic envelope of the dimeric species. Hence, the mass of the tetramers must be less than twice the mass of the dimers. Thus it is established that the tetrameric complex contains multiple covalent links, presumably between the S100A4 subunits.

In order to hypothesize about the biological significance of the different types of dimerization of S100A4 that were observed, comparison with S100B, which also forms disulphide bonds, is beneficial. Growing evidence indicates that members of the S100 protein family have extracellular roles [13]. For S100B protein it has been shown that its disulphide-bonded dimer is associated with a number of extracellular activities contributing to neuronal development, differentiation and brain repair [14–16]. Oxidized dimers might be easily formed in the extracellular environment. However, available data suggest that only a very small amount of disulphide dimer is produced. Although the dimer is a minor species, it may be capable of producing a strong biological effect. S100A4 has four Cys residues and it has been demonstrated that at least two different types of dimers are formed (I. Bronstein, unpublished work), and these dimers have different activities toward S100A4 targets. In addition, oligomeric forms of S100A4 are able to stimulate neurite outgrowth activity [18]. The number of oxidized dimers and oligomers might be increased in the presence of copper [19]. It has been found that copper stimulates oxidation of non-covalent S100B dimers. The sensitive detection by ESI of S100 proteins oxidized dimers in serum and various biological fluids might provide a powerful technique to study pathological conditions and also to monitor the formation of oxidized species in vitro for pharmacological purposes.

Scotto et al. [20] have shown that oxidation of the S100B cysteine residues, Cys-68 and Cys-84, induces a conformational change in the protein structure, unmasking a canonical CKII phosphorylation site located within the typical EF-hand calcium-binding site IIβ. In the case of S100A4 we have demonstrated by ESI FTICR that when S100A4 dimers are covalently linked through oxidized cysteine residues there are conformational changes compared with the non-covalently bound species.

Conclusions

Results have been presented for S100A4 demonstrating how non-covalently bound complexes can be distinguished unambiguously from covalently bound species. It has been shown that for unambiguous identification of a given peak in a mass spectrum the number of charges associated with that peak must be known. If the mass resolution obtained is not sufficient to resolve the isotopomers within a given charge state then the charge associated with that peak must be obtained by other indirect means and hence the mass cannot be known with absolute certainty. Additionally any peak within the spectrum that consists of an overlap of two isotopic distributions and the isotopomers are not resolved, for example a mixture of monomer and covalent dimers, then the mass measured will be that of neither the monomer nor the covalent species. If the resolution is high enough to allow isotopic resolution of the species then the reality that the peak is composite is apparent.

It has been demonstrated that a single mass spectrum of S100A4 can under the correct conditions reveal much more about the protein than simply the molecular mass. It has been established that the protein was present as monomers, both covalent and non-covalent dimers and as covalently bound tetramers. Additionally it has been shown that covalent and non-covalent dimers of S100A4 exist in different conformations, and comparison with S100B allowed us to hypothesize about the biological roles of the two forms of S100A4 dimer observed.

FTICR MS affords mass resolution high enough that the isotopomers of proteins up to 100 kDa can be resolved. For this reason FTICR MS is the only technique that allows the unambiguous identification of the masses of such proteins.

References

17. Reference deleted.

Received 20 June 2003