University of Alberta

Carbohydrate-Protein Interactions Studied Using Electrospray Ionization Mass Spectrometry

by

Hong Lin

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Chemistry

Fall 2013
Edmonton, Alberta

Abstract

This thesis describes the studying of electrospray ionization (ESI) process and the application of the direct ESI-MS assay to study noncovalent carbohydrate-protein interactions in vitro.

Protein unfolding induced by ESI process was first investigated. It is proposed that the Coulombic repulsion between the negatively charged residues and liquid/droplet surface charge under certain ESI conditions can induce unfolding of acidic proteins. A deleterious non-uniform response factors phenomenon induced by high molecular weight molecules and complexes in the application of direct ESI-MS assay was also investigated. It is possibly due to the reduction in the number of available surface sites in the ESI droplets upon introduction of large solute and increased competition between protein and the more hydrophilic carbohydrate-protein complex for these sites.

Direct ESI-MS assay was also utilized to investigate the stepwise binding of the GM1 pentasaccharide (GM1os) to the cholera toxin B subunit homopentamer and to elucidate positive binding cooperativity.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Ab	Abundance
ACS	Average charge state
B	Magnetic field
BCA	Bicinchoninic acid
CEM	Chain ejection model
CRM	Charge residue model
CSD	Charge state distribution
CT	Cholera toxin
CTB_5	Cholera toxin B subunit homopentamer
Da	Dalton
DC	Direct current
DDM	*n*-dodecyl-β-D-maltoside
DMPC	1,2-dimyristoyl-*sn*-glycero-3-phosphocholine
ELISA	Enzyme-linked immunosorbent assay
ESI	Electrospray ionization
ESI-MS	Electrospray ionization mass spectrometry
FT	Fourier transform
FTICR	Fourier transform ion cyclotron resonance
GM1	monosialotetrahexosylganglioside
GM1os	β-D-Gal*p*-(1→3)-β-D-Gal*p*NAc-(1→4)[α-D-Neu5Ac-(2→3)]-β-D-Gal*p*-(1→4)-β-D-Glc*p*

GSL	Glycosphingolipids
H-bond	Hydrogen bond
IEM	Ion evaporation model
IMS	Ion mobility separation
IS	Internal standard
ITC	Isothermal titration calorimetry
K_a	Association constant
$K_{a,\,q}$	Association constant for stepwise ligand binding
$K_{a,\,int}$	Intrinsic association constant
K_d	Dissociation constant
L	Ligand
L_{proxy}	Proxy ligand
L_{ref}	Reference ligand
MD	Molecular dynamics
MS	Mass spectrometry
MW	Molecular weight
m/z	Mass-to-charge ratio
nanoESI	Nanoflow electrospray ionization
ND	Nanodiscs
NOESY	Nuclear overhauser effect spectroscopy
NMR	Nuclear magnetic resonance
O	Octasaccharide [α-D-Gal*p*-(1→2)-[α–D-Abe*p*-(1→3)]-α-D-Man*p*-(1→4)-α-L-Rha*p*]$_2$

P	Protein
P_{ref}	Reference protein
R	The ratio of total abundance
RF	Radio frequency
RF	Response factor
SAS	Solvent accessible surface
scFv	Single chain antibody of the monoclonal antibody Se 155-4
S/N	Signal-to-noise ratio
SPR	Surface Plasmon resonance
T	Trisaccharide Methylα-D-Tal*p*-(1$\rightarrow$2)-[α-D-Abe*p*-(1$\rightarrow$3)]-α-Man*p*]
TM	Trajectory method
TcdB	Clostridium difficile toxin B
TOF	Time of flight
TSP	Tailspike protein
ω_c	Cyclotron frequency

Chapter 1

Carbohydrate-Protein Interactions Studied Using Electrospray Ionization Mass Spectrometry

1.1 Introduction

Carbohydrates are important and the most abundant biological molecules.[1-2] They are commonly found in the form of glycopeptides, glycoproteins and glycolipids at the cell surface, where they are positioned to interact with suitable proteinaceous receptors, such as lectins, antibodies and carbohydrate-processing enzymes in solution or the surfaces of other cells.[3] Carbohydrate-protein interactions play critical roles in a wide range of physiological and pathological cell functions, such as inflammation, fertilization, cell-cell and cell-matrix interactions, signal transduction, infections by microbes and the immune response.[4-6] The interactions between carbohydrates and proteins are driven primarily by the formation of hydrogen bond (H-bond) networks and van der Waals contacts.[7] Due to the low affinities that are typical of individual carbohydrate-protein interactions ($K_a < 10^4\ M^{-1}$),[8-10] many carbohydrate-binding proteins and protein complexes possess multiple carbohydrate binding sites and exploit multivalent binding to achieve high avidities.[11-12]

Investigations into the structure and thermodynamics and kinetics of carbohydrate-protein interactions *in vitro* are both of fundamental importance and serve to improve disease diagnosis and guide the development of new therapeutics. There are a number of established analytical methods available to identify and

quantify carbohydrate-protein interactions *in vitro*, each with particular strengths and weaknesses. Among the most widely used methods for measuring the association constants (K_a) are isothermal titration calorimetry (ITC),[13-14] surface plasmon resonance (SPR) spectroscopy,[15] enzyme-linked immunosorbent assay (ELISA),[16] and nuclear magnetic resonance (NMR) spectroscopy.[17]

ITC is a powerful and one of the most important techniques, which is generally considered as the "gold standard" for the study of thermodynamics of biological complexes like carbohydrate-protein interactions. It is the only method that can provide association constant (K_a), Gibbs free energy of binding, enthalpy and entropy directly and accurately from a single experiment.[18-19] Conventional ITC instruments suffer from low sensitivity and generally require large amounts (~mg) of pure protein and ligand for each analysis. However, new ITC technologies, such as the Nano ITC™, have improved sensitivity and substantially lowered sample requirements.[20]

SPR spectroscopy represents another widely used method for evaluating both the kinetic[21-22] and thermodynamic parameters[23-25] of carbohydrate-protein interactions. This technique affords high sensitivity, and requires only a very small amount of sample (~ng) for each measurement. A potential limitation of this approach is the need to immobilize one of the binding partners (usually the ligand) on a sensor chip, which may affect the nature of the binding interaction.

ELISA is also a widely used method for quantifying carbohydrate-protein interactions.[16] While there are several different ways of implementing ELISA, the assay requires the immobilization of one of the binding partners on the solid phase

followed with incubation with solutions containing the other binding partner which is attached with an enzyme to help generate a signal that can be properly quantified. This method, once setup, is fast and relatively sensitive. However, the relatively large immobilization surface area can lead to nonspecific binding and increased background. Also, most ELISAs rely upon enzyme-mediated amplification of signal to achieve reasonable sensitivity, which can limit its applicability.

NMR is also widely used to characterize the structures of biological molecules and their complexes in solution and to quantify the strength of the interactions.[26-29] Recently, transferred NOSEY NMR spectroscopy has been used to estimate the dissociation kinetics for streptavidin-small ligand.[30] However, NMR measurements are usually limited to relatively small proteins, with molecular weights (MWs) <40 kDa.[26] Additionally, NMR measurements generally require large amounts of sample (typically mg quantities) and are time consuming process, which limit its application.

Recently, electrospray ionization mass spectrometry (ESI-MS), has emerged as an important tool for identifying and quantifying (stoichiometry and affinity) carbohydrate-protein interactions, as well as other non-covalent protein-ligand complexes in vitro.[31-37] The direct ESI-MS assay, which is the main focus of this thesis, relies on the direct detection and analysis of the protein or protein-ligand complex ions in the gas phase. The first quantitative study of carbohydrate-protein binding using the direct ESI-MS assay was reported by Kitova *et al.* in 2001 and involved the weak interactions between analogues of the P^k trisaccharide and B_5

homopentamer of Shiga-like toxin type I.[38] The K_a values measured by ESI-MS were found to be in a good agreement with the values reported by ITC.[39] Since then, the direct ESI-MS assay has been used to quantify carbohydrate interactions with a wide variety of carbohydrate-binding proteins (antibodies, bacterial toxins, lectins and carbohydrate-processing enzymes).[36-37, 40-44]

The direct ESI-MS assay possesses a number of advantages, including its simplicity (labeling or immobilization free), speed (individual K_a measurements normally can be completed in less than a few minutes). Additionally, when performed using nanoflow ESI (nanoESI), which can be operated at solution flow rate in the nL/min range, this assay typically consumes only pmol or even less of analyte per analysis. More importantly, the ESI-MS assay can afford direct establishment of the binding stoichiometry and the ability to measure multiple binding equilibria directly and simultaneously. These features enable the determination of both the macroscopic and microscopic K_a values for stepwise binding models involving multisubunit protein, even being suitable to elucidate cooperative binding at the same time.[45-48] This assay also naturally lends itself to monitoring and quantifying protein-ligand interactions in solutions containing mixtures of ligands and/or proteins.[49-52]

However, like all other techniques, the ESI-MS assay has its limitations. An important underlying assumption in the implementation of this assay is that the interactions and stiochiometries of protein-ligand complex existed in the solution, can be exactly transferred into gas phase without any change. However, it's not the case all the time. The potential pitfalls of this assay will be discussed in more

detail later on. Although the ESI process is not fully understood yet, the ESI-MS binding assay has become a promising and most powerful technique for characterizing carbohydrate-protein interactions.

Before describing the strategies of ESI-MS assays in detail, it is necessary to first review the basic principles of ESI. An overview of ESI mechanism and followed by the instrumentation, is given below.

1.2 Electrospray ionization (ESI) mass spectrometry

1.2.1 ESI mechanism

ESI is a versatile and soft ionization technique allowing proteins or protein-ligand complexes to be transferred into the gas phase as intact ions (Figure 1.1). It occurs at atmospheric pressure.[53] The mechanism of the ESI process, as described by Kebarle and coworkers,[54] involves three major steps:

a) Production of charged droplets at the ESI capillary tip.

b) Shrinkage of the charged droplets due to solvent evaporation and repeated charge induced droplet disintegrations.

c) The production of the gas-phase ions from these droplets.

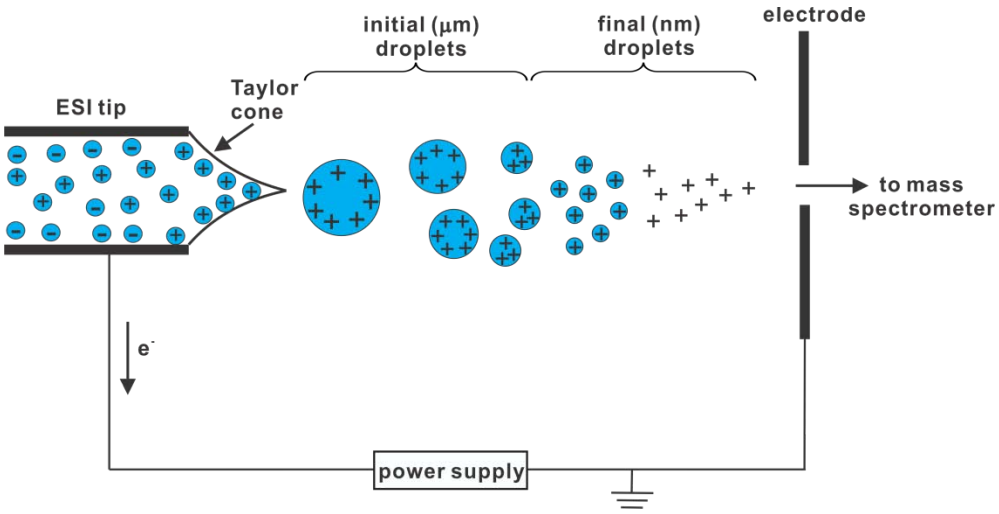


Figure 1.1[*] Schematic representation of ESI carried out in positive ion mode and the processes that lead to the formation of gas phase ions.

a) Production of charged droplets at the ESI capillary tip:

Shown in Figure 1.1 is a schematic diagram of ESI in positive ion mode. The high positive voltage applied to the capillary induces charge separation of electrolytes in solution. The positive charges drift towards the liquid surface leading to the formation of a liquid cone referred to as a Taylor cone.[55] The increase of surface area due to the cone formation is resisted by the surface tension of the liquid. Under sufficiently high field, the liquid cone becomes unstable that emits a fine mist of droplets.[55] This spraying process is usually assisted by a coaxial gas flow (not shown in Figure 1.1).[53] The initial ESI droplets usually have radii in the micrometer range.[54]

b) Shrinkage of the charged droplets due to solvent evaporation and repeated charge induced droplet disintegrations:

[*] Modified from reference 59

The charged droplets produced at the spray tip will shrink due to rapid solvent evaporation causing an increase in the electric field normal to the surface of the droplets while the charge remains constant. The energy required for the solvent evaporation is provided by the thermal energy of the ambient gas, air at atmospheric pressure in most cases. As the droplets get smaller and the charge density on the shrinking droplets builds up until the Rayleigh limit, the point at which the Coulombic repulsion of the surface charges is balanced by the surface tension of the droplets,[56] This leads to fission of the droplets that typically releases a jet of smaller and highly charged progeny droplets. Repeated evaporation/fission events ultimately yield the final generation of ESI droplets with radii of a few nanometers.

c) The production of the gas-phase ions from these droplets:

Three different models (IEM, CRM, and CEM)[31, 54, 57-59] have been proposed to account for the formation of gas-phase ions from the very small and highly charged droplets.

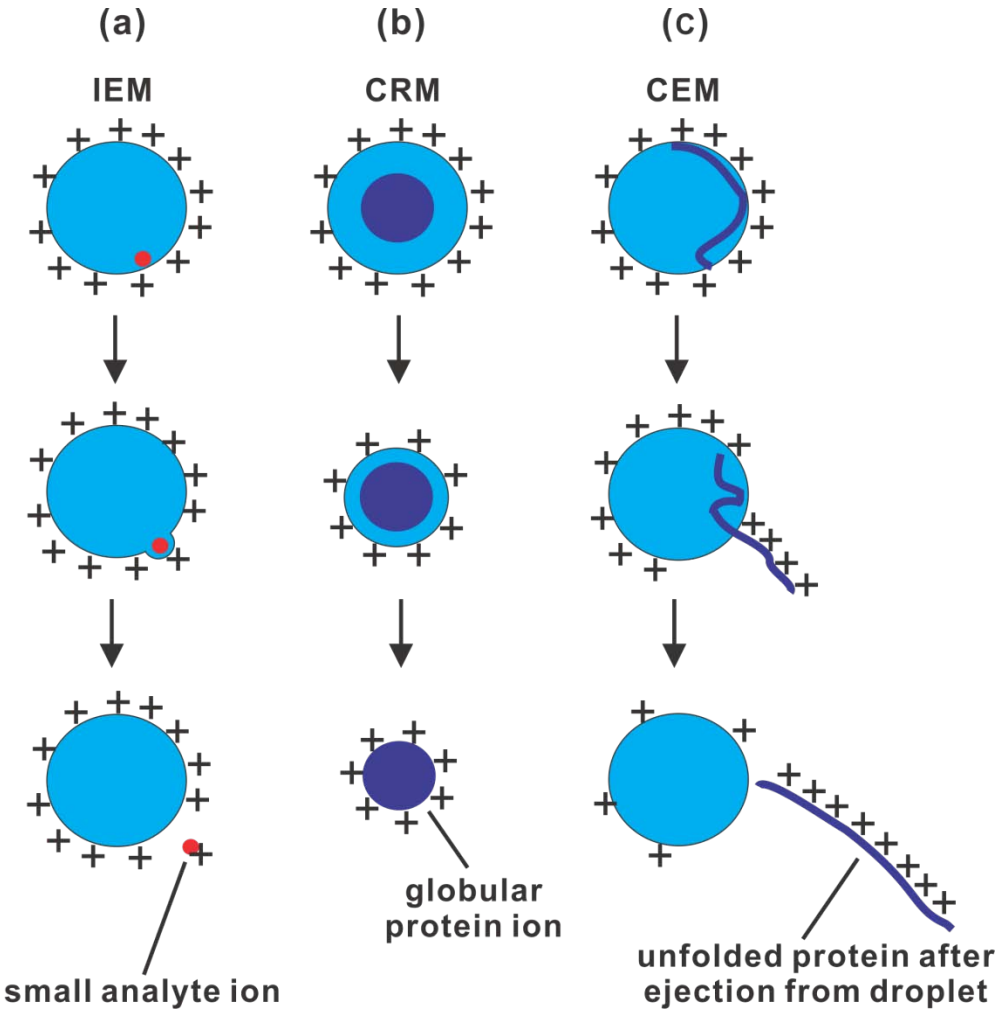


Figure 1.2[*] Summary of different models for the formation of gas phase ions. (a) IEM: Small ion ejection from a charged nanodroplet. (b) CRM: Formation of a globular protein into the gas phase. (c) CEM: Ejection of an unfolded protein.

i) The ion evaporation model (IEM):[60] This model (Figure 1.2a), proposed by Iribarne and Thomson,[61] predicts that direct ion emission from the droplets will occur after the radii of the droplets shrink to radii less than 10 nm. The IEM model is experimentally well-supported for small (in)organic ions.

[*] Adapted from reference 59

ii) The charged residue model (CRM):[62] This model was proposed by Dole[62] and coworkers. It is widely accepted that gas phase ions of large globular species such as natively folded proteins are formed via this model (Figure1.2b). For the CRM, Rayleigh-charged nanodroplets that contain a single analyte evaporate to dryness. As the last solvent shell disappears, the charge of the vanishing droplet is transferred to the analyte,[54, 63] leading to the formation of monomodal and narrow charge state distributions (CSDs) centered at high m/z values in the corresponding ESI-MS spectrum. CRM nanodroplets remain close to the Rayleigh limit throughout the entire shrinkage process, implying that the droplet sheds charge as its radius decreases.[59]

iii) The chain ejection model (CEM):[64-65] MD simulations revealed that unfolded proteins undergo ESI via this model (Figure 1.2c). Protein unfolding in solution can be triggered, such as by exposure to an acidic solvent. The resulting conformers are highly disordered, and are now solvent accessible, which switches the properties of compact/hydrophilic protein to extended/hydrophobic.[66] In a Rayleigh-charged nanodroplet, unfolded chains immediately migrate to the droplet surface. One chain terminus then gets expelled into the vapor phase. This is followed by stepwise sequential ejection of the remaining protein residues and separation from the droplet.[59] As reflected in the spectrum, unfolded proteins produce a wider charge-state envelop centered at much lower m/z values.

Recently Michael Gross and co-workers[67] propose a modification of CRM in which CRM is preceded by IEM. This mechanism is expected to operate when

salt additives (buffers) such as ammonium acetate or triethylacetate are present in millimolar concentrations in the solution that is electrosprayed.

In the present work, nanoESI was used in all cases. The mechanism of nanoESI is the same as that of ESI, except that by using a narrow glass tip, nanoESI operates at lower solution flow rates (10-50 nL/min) than conventional ESI (1-10 μL/min) and correspondingly emits smaller droplets.[68] Thus, only picomoles or less of analyte is required per analysis, a very important feature in the analysis of biological molecules where there is often only a limited amount of sample.[69] In addition to being more sensitive than ESI, nanoESI readily allows the transfer of noncovalent complexes from buffered aqueous solutions to the gas phase and, therefore, can be directly performed on complex solutions that more closely resemble physiological conditions.[70] The short lifetimes of the nanoESI droplets, which are estimated to be tens of microseconds, are also likely advantageous for preserving the original solution composition throughout the formation of gaseous ions. Furthermore, nanoESI can minimize nonspecific aggregation that can occur during ESI process as there are fewer analyte molecules per droplet.[71-72] These features of nanoESI make it the method of choice for investigating noncovalent complexes directly by MS.

1.2.2 MS instrumentation

There are a number of different types of mass analyzers, including ion trap, quadrupole, magnetic sector, Fourier transform ion cyclotron resonance (FTICR) and time of flight (TOF). In the present study, nanoESI combined with FTICR

and hybrid quadrupole-ion mobility separation (IMS)-TOF mass spectrometers were used.

1.2.2.1 Fourier transform ion cyclotron resonance (FTICR) mass spectrometer

Shown in Figure 1.3a is a schematic diagram of the Bruker Apex-II nanoESI-FTICR mass spectrometer used in the present work (Chapter 2). Gaseous ions are produced by nanoESI performed at atmospheric pressure from buffered aqueous solutions containing analytes by applying a high voltage (typically +(-)1000 V) to a platinum (Pt) wire inserted into the solution in the glass tip. Small droplets produced by nanoESI are sampled into the mass spectrometer through a heated metal capillary, and gaseous ions are transmitted through a skimmer and accumulated in the hexapole for certain time period to enhance the signal-to-noise (S/N) ratio. After accumulation, ions are ejected from the hexapole, accelerated by a high voltage through the fringing field of a 9.4T superconducting magnet, decelerated, and trapped by a combination of electric and magnetic field in FT-ICR cell for detection. The typical base pressure for the instrument is ~5×10^{-10} mbar, maintained by the differential pumping system.

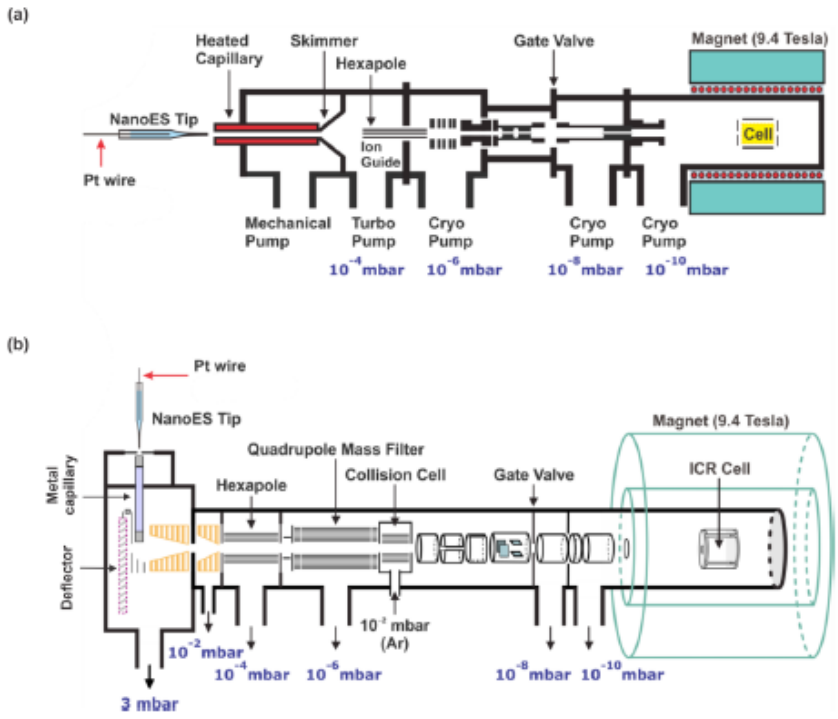


Figure 1.3 Schematic diagrams of (a) the Bruker Apex-II 9.4T nanoESI-FTICR mass spectrometer and (b) the Bruker Apex-Qe 9.4T nanoESI-FTICR mass spectrometer used in this study. Figures were reported reproduced from the Bruker user's manual.

The other FTICR mass spectrometer used in this thesis (Chapters 3 and 4) is a Bruker Apex-Qe nanoESI-FTICR mass spectrometer (Figure 1.3b). It is a hybrid quadrupole-FTICR mass spectrometer, in which two mass analyzers are combined. The quadrupole can act as a mass filter to select and isolate targeted analyte ions efficiently for tandam MS (MS/MS) analysis. The operation scheme of the Apex-Qe is very similar to Apex-II, however the ion source represents the main difference. The ions generated in the electrospray process, with the assistance of a neublizer and counter-drying gas enter the vacuum system of the Apex-Qe

through a metal capillary. From the capillary exit, ions enter the transmitting funnels and skimmers orthogonally, with the assistance of a deflector. The ions are then stored electrodynamically in the hexapole followed by further accumulation in the quadrupole. In the present work, the quadrupole was operated in radio frequency (RF)-only mode as it acted as a wide band-pass filter to transmit ions for further analysis. After accumulation, the ions are transferred through a series of ion optics into the ICR cell for detection.

FTICR mass spectrometers were used in this study for its high resolving power and mass accuracy. The general operating principles of FTICR are described in many reviews[73-75] and, therefore, only a brief overview is given here. The ICR cell consists of three pairs of plates (trapping, excitation and detection) and is located inside a spatial uniform static superconducting high field magnet cooled by liquid helium and liquid nitrogen. When the ions pass into the magnetic field they are bent into a circular motion in a plane perpendicular to the field (see Figure 1.4) by the Lorentz force. The cyclotron frequency, ω_c is expressed in eq. 1.1:

$$\omega_c = \frac{qB}{m} = \frac{z\,eB}{m} \qquad (1.1)$$

where ω_c is the cyclotron frequency, q is the charge of the ion ($q = ze$, where z and e are the charge and the elementary charge, respectively), B is the magnetic field strength and m is the mass of the ion. To obtain the cyclotron frequency in Hertz (f) the results in radian per second has to be divided by 2π (i.e. $\omega_c = 2\pi f$). A notable feature of equation 1.1 is that all ions of a given m/z rotate at the same frequency, independent of their velocities. The ultrahigh resolution achieved by

FTICR MS is a direct result of insensitivity of the cyclotron frequency to the kinetic energy of an ion.

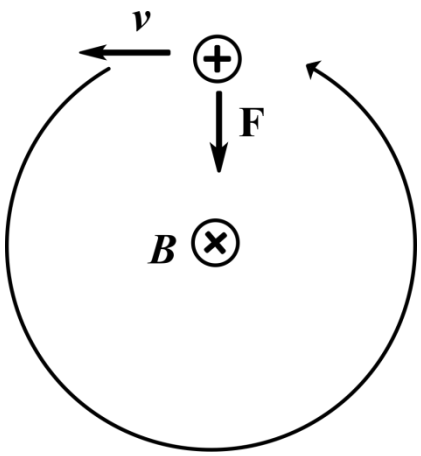


Figure 1.4 Illustration of the cyclotron motion of a positive ion of charge q moving at velocity $\boldsymbol{v}$ in the presence of a constant magnetic field, $\boldsymbol{B}$, which is pointing orthogonal to the motion of the ion. The ion moving to the left experiences a downward Lorentz force, $\mathbf{F} = q\ (\boldsymbol{v} \times \boldsymbol{B})$, $q = ze$, resulting in a counterclockwise orbit.

Ions moving in cyclotron orbits in a static magnetic field will not generate much signal if placed between a pair of detection electrodes. In order to produce a measurable signal for the ions trapped in the ICR cell to be detected, a packet of ions of a given *m/z* needs to be excited by applying an oscillating electrical field such as provided by an AC signal generator. If the frequency of the applied field is the same as the ω_c of the ions, the ions will absorb energy and thus increase their orbital radius but keep a constant cyclotron frequency. Shown in Figure 1.4

is the spiral trajectory of the excited ions with the same *m/z* and ω_c. As the coherently orbiting excited ions passing another opposing pair of electrodes (detection plates) of the cell, also parallel to the magnetic axis, they induce an alternating current to the plates called image current (Figure 1.5). The amplitude of this image current is proportional to the number of ions in the analyzer ICR cell, while the frequency of the alternating current matches the cyclotron frequency of ions. FT transforms the detected image current from the time domain signal into the frequency domain and a mass spectrum can be generated as ω_c is related to *m/z*. Once this transient signal is amplified and detected, the ions are detected without colliding with the electrodes, which makes the detection scheme non-destructive and allows for the improved sensitivity and versatility of FTICR.

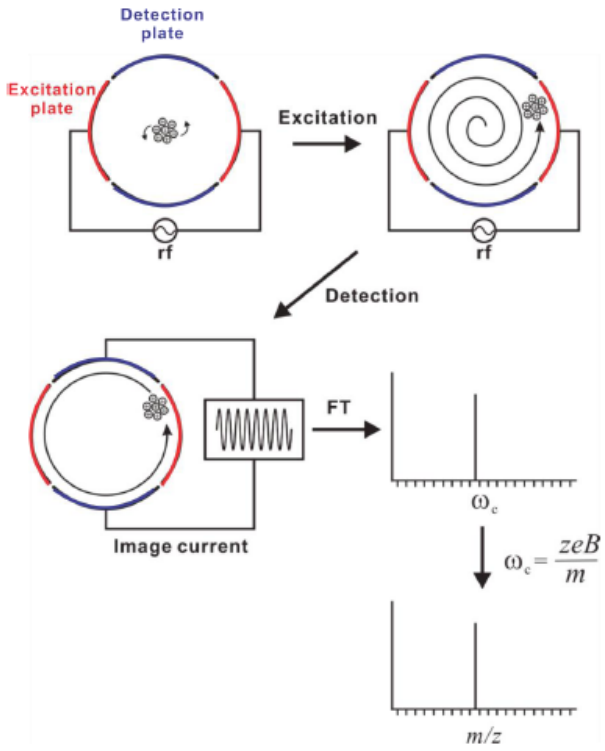


Figure 1.5[*] Illustration of excitation, image current detection and the production of mass spectrum by FTICR.

A Fourier transform transforms the detected image current into a frequency domain from the time domain signal and a mass spectrum can be registered because the cyclotron frequency is related to m/z (equation 1.1). As the cyclotron frequency can be measured with very high precision, the mass accuracy of FTICR MS is as high as 1 ppm. The resolving power of FTICR MS can routinely reach hundreds of thousands at broad band mode, typically measured as the full width at half maximum (FWHM). The resolving power is proportional to the magnetic field strength (with higher magnetic field having higher resolution), and the

[*] Adapted from Bruker Daltonics

acquisition time. The acquisition time is the duration of the detection phase, determined by the dataset size and the frequency of sampling. Longer acquisition time (larger dataset size) results in higher resolution in spectrum. Also, high vacuum (10^{-10} mbar) is necessary in the cell region of FTICR MS, to avoid the collision with gas particles and the deactivation of the ions.

1.2.2.2 Hybrid Quadrupole Time of Flight mass spectrometer

A Synapt G2 quadrupole ion-mobility separation time-of-flight (Q-IMS-TOF) mass spectrometer (Waters UK Ltd., Manchester, UK), equipped with a nanoflow ESI (nanoESI) source was used in this work (Chapter 2 & 3) (Figure 1.6).

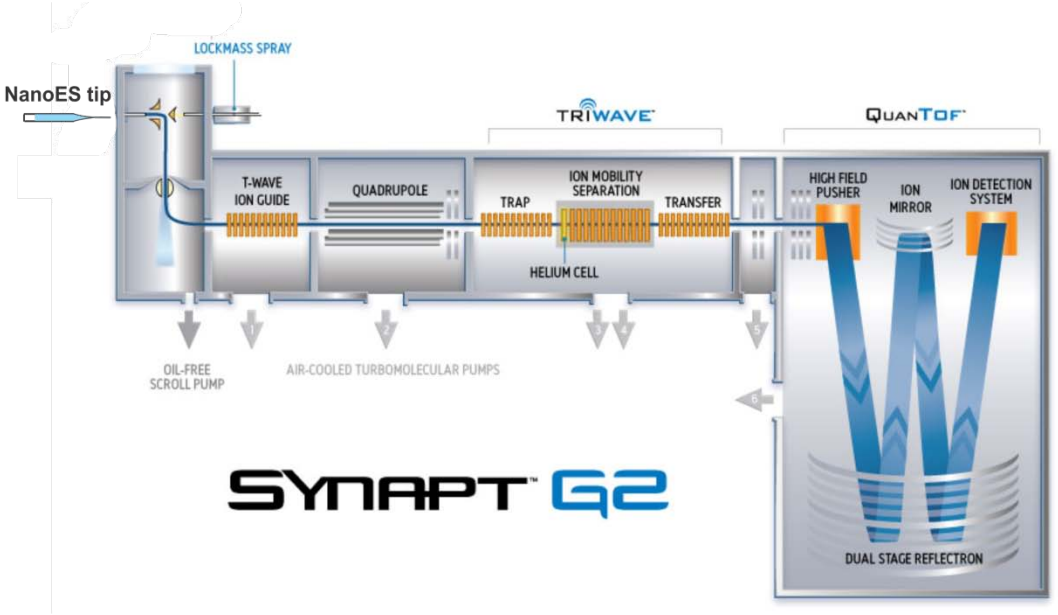


Figure 1.6[*] Schematic representation of the Waters Synapt G2 nanoESI-quadrupole-IMS-TOF mass spectrometer used in this study.

[*] Adapted from http://www.waters.com

Briefly, fine droplets produced by nanoESI are introduced into the mass spectrometer through a "Z-spray source", which minimizes neutral contamination and enhances the signal-to-noise. The nanoESI equipped with this mass spectrometer is similar to the one mentioned previously. The resulting gaseous ions are then transmitted through a quadrupole mass filter to the ion mobility section of the instrument (Triwave). The mobility separated ions are then detected by an orthogonal acceleration (oa)-TOF mass analyzer (QuanTOF^{TM}) equipped with a high field pusher and a dual-stage reflectron.

The Waters Synapt G2 nanoESI-quadrupole-IMS-TOF mass spectrometer was used in this study for its wide mass range and high sensitivity. The IMS feature of this mass spectrometer is actually not used in this study, as a result, only a brief overview of the quadrupole and TOF parts of the instrument is given following.

1.2.2.2.1 Quadrupole

The quadrupole is composed of four cylindrical metal rods that are accurately positioned in a radial array and the diametrically opposed rods are paired. A direct current (DC) potential and a radiofrequency (RF) potential, 180 degrees out of phase, are applied to each pair of rods.[75] Depending on the specific voltage and frequency applied, ions of a particular m/z ratio can be selected and transit down the entire length of the rods, while other ions outside the m/z range hit the rods and are expelled. The quadrupole can also act as a broad bandpass filter, by turning off the DC voltages and operating in RF only mode, that

transmits and guides ions over a wide mass-to-charge (m/z) range to other components of the apparatus. In the Synapt mass spectrometer, the quadrupole contains two parts, a quadrupole prefilter followed by a quadrupole mass filter (Figure 1.7). The use of prefilter increases the absolute sensitivity by minimizing the effects of fringing fields at the entrance to the quadrupole.[76]

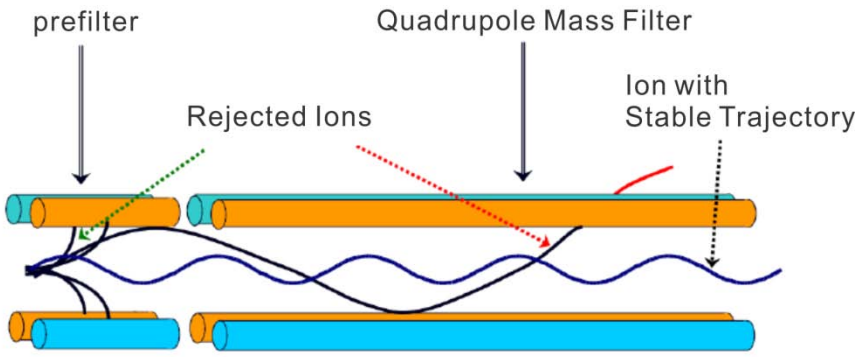


Figure 1.7 Schematic representation of the quadrupole in Waters Synapt G2 mass spectrometer.

1.2.2.2.2 Time of Flight (TOF)

For TOF analyzers, the physical property that is measured during an analysis is the flight time of the ions.[77] mass-to-charge (*m/z*) values are determined by measuring the time that ions take to move through a field-free region (flight tube) between the source and the detector, according to eq 1.2:

$$(m/z)^{\frac{1}{2}} = t(\frac{\sqrt{2eV_s}}{L}) \qquad (1.2)$$

where *m* is the mass of the ion, *z* is the charge state of the ion, *e* is the elementary charge, V_s is the acceleration potential, t is the flight time and L is the length of the flight tube. This equation shows that *m/z* can be calculated from a measurement of t, the terms in parentheses being constant. The lower the mass of an ion, the faster it will reach the detector. There are two types of TOF analyzers: linear TOF analyzer and reflectron TOF analyzer. The linear TOF analyzer has the drawback that ions of the same *m/z* may reach the detector at different times, due to initial energy distribution, resulting in peak broadening and poor resolution.

In Waters Synapt G2 mass spectrometer, a reflectron TOF analyzer is used. The reflectron TOF analyzer compensates the energy distribution of ions by using successive sets of electric grids of increasing potential which deflects the ions and reverses their flight direction sending them back through the flight tube. Depending on their kinetic energy, ions of the same *m/z* will penetrate the field at different depths; ions with more kinetic energy and hence with faster velocity will penetrate the field more deeply than ions with lower kinetic energy. Consequently, the faster ions will spend more time in the reflectron and will reach the detector at the same time as the slower ions with the same *m/z*. The net effect is improved mass resolution typically in the range of 10,000 – 20,000 with minimal losses in sensitivity.

After introducing the review of the basic principles and instrumentations of ESI-MS, a more detailed introduction of ESI-MS binding assays for studying noncovalent carbohydrate-protein interactions is given in the following section.

1.3 Direct ESI-MS binding assay

The direct ESI-MS binding assay is based on the direct detection of free and ligand-bound protein ions by ESI-MS. For a solution containing protein P, and ligand L, when the equilibrium is reached,

$$P + L \rightleftharpoons PL \tag{1.3}$$

the association constant (K_a) is expressed as

$$K_a = \frac{[PL]_{eq}}{[P]_{eq}[L]_{eq}} \tag{1.4}$$

The equilibrium concentrations, $[PL]_{eq}$, $[P]_{eq}$ and $[L]_{eq}$, are calculated from the initial concentrations of protein and ligand in solution, $[P]_o$ and $[L]_o$, and the relative abundance of the corresponding bound and unbound protein ions measured in the mass spectrum (eq. 1.5-1.7). The ratio (*R*) of the total abundance (*Ab*) of bound and unbound protein ions (e.g., PL^{n+}, P^{n+}) measured in the gas phase by ESI-MS is expected to be equivalent to the ratio of the concentrations in solution at equilibrium.[78]

$$R = \frac{Ab(PL)}{Ab(P)} = \frac{[PL]_{eq}}{[P]_{eq}} \tag{1.5}$$

$$[P]_o = [P]_{eq} + [PL]_{eq} \tag{1.6a}$$

$$[L]_o = [L]_{eq} + [PL]_{eq} \tag{1.6b}$$

$$[PL]_{eq} = \frac{R[P]_o}{1+R} \tag{1.7}$$

Then K_a value for the 1:1 protein-ligand complex is determined from the ratio *R*, $[P]_o$ and $[L]_o$, eq 1.8.

$$K_a = \frac{R}{[L]_o - \frac{R}{1+R}[P]_o} \quad (1.8)$$

When the protein (or protein assembly) can bind to Q ligands (where $Q > 1$), there are Q reactions to be considered:

$$P + L \rightleftharpoons PL \quad (1.9a)$$

$$PL + L \rightleftharpoons PL_2 \quad (1.9b)$$

$$PL_2 + L \rightleftharpoons PL_3 \quad (1.9c)$$

……

$$PL_{Q-1} + L \rightleftharpoons PL_Q \quad (1.9d)$$

Here, we only describe the simplest case, in which all Q binding sites are equivalent, with identical intrinsic association constants $K_{a,int}$. The treatment of more complicated cases has been discussed elsewhere.[79] The equilibrium concentrations, $[P]_{eq}$, $[PL]_{eq}$, …, $[PL_Q]_{eq}$, can be determined from relative abundance of the corresponding ions observed in the mass spectrum and eq 1.10a. Then, using these values, the equilibrium concentrations of L can be found from eq 1.10b:

$$[P]_o = [P]_{eq} + [PL]_{eq} + [PL_2]_{eq} + \cdots + [PL_Q]_{eq} \quad (1.10a)$$

$$[L]_o = [L]_{eq} + [PL]_{eq} + 2[PL_2]_{eq} + \cdots + Q[PL_Q]_{eq} \quad (1.10b)$$

$K_{a,q}$ (apparent association constant for formation of protein bound with q ligands) can be determined from eq 1.11, which are based on the general expression $K_{a,q} = K_{a,int}(Q - q + 1)/q$, where q is the number of occupied binding sites:[79]

$$K_{a,q} = \frac{[PL_q]_{eq}}{[PL_{(q-1)}]_{eq}[L]_{eq}} \quad (1.11)$$

In practice, ESI-MS binding measurements are usually limited to *R* values ranging from approximately 0.05 to 20 and P and L concentrations in the 0.1 to 1000 μM range. It follows that K_a values accessible with the direct ESI-MS binding assay range from approximately 10^3 to 10^7 M^{-1}.[80] However, interactions with much larger K_a values can be probed using competitive binding and direct ESI-MS measurements.

1.4 Potential pitfalls of the direct ESI-MS binding assay

As mentioned before in session 1.1, the direct ESI-MS has certain limitations. Any physical or chemical process that alters the equilibrium abundance ratio of bound-to-free protein during ESI process and in the gas phase from that present in bulk solution will lead to incorrect K_a and, potentially, obscure the true binding stoichiometry. Three common sources of error associated with the ESI-MS measurements are: (1) non-uniform response factors, (2) nonspecific binding and (3) in-source dissociation. Each of these sources of error is briefly described below, along with current strategies for minimizing their effects on the binding measurements.

1.4.1 Non-uniform response factors

As described above, the abundances of gaseous P and PL ions measured by ESI-MS are related to the solution concentrations by a response factors (*RF*s), which collectively accounts for the ionization and detection efficiencies, eq 1.12:

$$\frac{[\mathrm{PL}]_{eq}}{[\mathrm{P}]_{eq}} = \frac{Ab(\mathrm{PL})/\,RF_{\mathrm{PL}}}{Ab(\mathrm{P})/\,RF_{\mathrm{P}}} = RF_{\mathrm{P/PL}}\,\frac{Ab(\mathrm{PL})}{Ab(\mathrm{P})} \qquad (1.12)$$

where RF_{P} and RF_{PL} are the response factors of P and PL, respectively, and $RF_{\mathrm{P/PL}}$ is the ratio of the corresponding *RF* values (referred to as the relative response factor). Although the absolute *RF* values depend on many factors – the size, structure and surface properties of P and PL, the solution composition and instrumental parameters used for the measurements - uniform *RF*s for P and PL (i.e., $RF_{\mathrm{P/PL}} \approx 1$) are expected in cases when the L is small compared to the P, such that the size and surface properties of the P and PL are similar.[36, 42-43, 81-83] Even though there is no firm guideline suggesting when this approximation is valid, it typically holds in cases where the molecular weight of PL and P (MW_{PL} and MW_P, respectively) are similar, i.e., $MW_{PL}/MW_P \leq 110\%$.[43] However, there are cases where the ESI-MS response of a protein complex is significantly different than the response of the free protein,[84-85] which is also investigated in the present work (Chapter 3). A variety of strategies have been developed to minimize the effects of non-uniform *RFs* on the determination of K_a values using ESI-MS assay. One approach involves the introduction of the $RF_{\mathrm{P/PL}}$ term as an adjustable parameter in an appropriate binding model, which is fit to the experimental data.[19, 84-87] However, this method requires fitting a model with multiple adjustable parameters to the titration data and, therefore, high quality

experimental data are required to obtain reliable K_a values.[86] Furthermore, there is another underlying assumption to use this approach, which is that $RF_{P/PL}$ is independent of concentration, at least over the range of the concentrations investigated. A variation on this method involves the use of an internal standard (IS). An appropriate IS is one with similar properties (MW and surface activity) to the analyte of interest, but which does not bind to L specifically.[87] The main advantage of this approach is that fluctuations in $RF_{P/PL}$ due to concentration, instability in the ESI or other factors, are reflected, at least to some extent, in the abundance of the IS ion. An alternative strategy involves monitoring the abundance of L, relative to that of an IS, as $[P]_0$ is varied.[88] In this assay, the IS resembles L but does not bind to P. The abundance ratio of L to IS ions serves to quantify the changes in [L] in solution as a function of $[P]_0$.[80]

1.4.2 Nonspecific binding

It is well established that, during the ESI process, free L can bind nonspecifically to P and PL (or PL_q in general) due to the concentration effects, resulting in false positives. Consequently, the observation of gaseous ions corresponding to a particular PL complex does not, by itself, establish the presence of that interaction in solution. The observation of multiple ligands bound to the target protein P with a Poisson-like distribution is a tell-tale sign of occurrence of nonspecific ligand binding. Also, changes in the magnitude of K_a with changes in ligand concentration may also alert to the occurrence of nonspecific ligand binding.

The formation of nonspecific protein-ligand (PL) complexes can be understood in the context of the CRM of ESI (section 1. 2.1). See Figure 1.8 for the formation during ESI process.

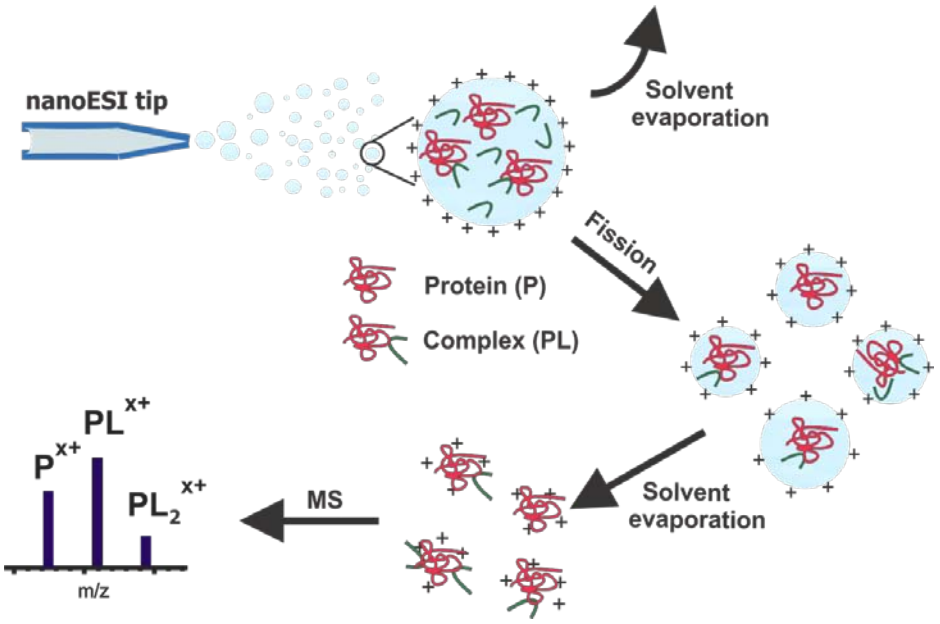


Figure 1.8[*] Cartoon of the CRM of ESI depicting the formation of nonspecific protein-ligand interactions (false positives).

According to the CRM, the initial ESI droplets undergo solvent evaporation until they come close to Rayleigh limit, at which point they undergo fission, releasing small multiply charged offspring nanodroplets containing no analyte or one or more molecules of analyte. Solvent evaporation from the nanodroplets ultimately yields gaseous ions. If a nanodroplet contains two or more analyte molecules, nonspecific intermolecular interactions can occur as the droplet evaporates to dryness, leading to the formation of nonspecific complexes.

[*] Adapted from reference 80

Nonspecific binding of L to P and PL obscures the true binding stoichiometry in solution and introduces error into the $K_{a,app}$ values measured by ESI-MS.

Generally, the formation of nonspecific protein–ligand complexes can be minimized by decreasing the initial concentration of ligand. For very weak ligand interactions ($K_{a,int} < 10^4$ M^{-1}), however, high initial concentrations of ligand (> 0.05 mM) are typically required to produce detectable level of complexes. In such cases, nonspecific binding is often unavoidable.

A number of strategies have been developed to correct ESI mass spectra for the occurrence of nonspecific binding.[89-94] The most straightforward approach is the reference protein method, which involves the addition of a reference protein (P_{ref}) that does not bind specifically to the protein and ligand of interest to the solution.[89] The method is based on the assumption that nonspecific ligand binding is a random process, as suggested by the observation that the distribution of nonspecifically bound molecules often resembles that of a Poisson process, and affects equally all protein species present in the ESI droplets. The occurrence of nonspecific protein-ligand binding is monitored by the appearance of ions corresponding to nonspecific complexes of P_{ref} and L in the mass spectrum.[89] The fraction abundance of P_{ref} undergoing nonspecific ligand binding provides a quantitative measure of the contribution of nonspecific binding to the measured intensities of protein and specific protein-ligand complexes. The "true" abundance of a given PL_q species can be calculated from the apparent (measured) abundance of the PL_q species ($Ab_{app}(PL_q)$) and the distribution of nonspecific $P_{ref}L_q$ species using the following expression:

$$Ab(PL_q) = [Ab_{app}(PL_q) - f_1,P_{ref}Ab(PL_{q-1}) - \cdots f_q,P_{ref}Ab(P)]/f_0,P_{ref} \qquad (1.13)$$

where f_q,P_{ref} is the fractional abundance of P_{ref} bound to q molecules of L. Notably, this method has been shown to successfully utilized to correct for the nonspecific binding of neutral, charged carbohydrates, amino acids, peptides and divalent metal ions to proteins during ESI-MS analysis.[89, 91, 93, 95]

1.4.3 In-source dissociation

Collision-induced dissociation of the gaseous PL complexes in the ion source can also alter the relative abundance of PL and P ions.[89] For a 1:1 PL complex, as a consequence, the *R* value mentioned previously is smaller; in-source dissociation will necessarily decrease the magnitude of K_a. In the extreme case, where no PL complex ions survive to detection, in-source dissociation results in a false negative.[41] The influence of in-source dissociation on binding measurements depends on the configuration of the ion source used, the choice of instrumental parameters as well as the gas-phase stability of the complex being investigated. Usually, the occurrence of in-source dissociation can be identified from changes in *R* resulting from changes in ion source parameters, especially voltage differences in regions of high pressure, that influence the internal energy of the ions.

In cases where the gaseous complexes are prone to in-source dissociation, low temperatures (sampling capillary, drying gas), low potentials across lens elements, and short accumulation times are essential for obtaining more reliable binding constants. However, these conditions normally reduce signal intensities

on the other hand. Thus, a balance must be found to minimize dissociation and obtain mass spectra with sufficient signal-to-noise ratio at the same time. Another way to minimize the in-source dissociation is through the addition of small organic molecules, such as imidazole, to the ESI solution.[89, 96] The origin of the stabilizing effects of imidazole is believed to be due, at least in part, to enhanced evaporative cooling resulting from the dissociation of nonspecific imidazole adducts from the gaseous PL ions.[89] Additionally, the use of imidazole, which has a relatively high gas phase basicity and a relatively low gas phase acidity,[97] may also lead to a reduction in the charge states of the protein complexions. Moreover, it was shown that the introduction of imidazole vapor to the ion source also protects complexes against in-source dissociation.[98] However, these approaches have its limitations as the detection of very labile gas phase complexes, which rapidly dissociate at ambient temperature, by ESI-MS remains challenging.

Recently new ESI-MS assay which is a competitive binding assay, such as the reference ligand method, was successfully developed to quantify protein-ligand interactions that are highly labile and cannot be detected by ESI-MS.[41] This assay employs the direct ESI-MS analysis in conjunction with a reference ligand (L_{ref}). The L_{ref} can bind specifically to P at the same binding site as L. The binding affinity between L_{ref} and P is known and they form a more stable protein-ligand complex in the gas phase. The fraction of P bound to L_{ref}, which can be determined directly from the ESI mass spectrum, is sensitive to the fraction of P bound to L in solution and enables the affinity of P for L to be determined. This method has proven particularly useful for the analysis of PL interactions that are

kinetically unstable in the gas phase at room temperature.[99]

1.4.4 Other sources of error

Lately evidence has emerged to show that the ESI process itself may induce conformational changes of protein as well as the ligand-protein complexes, further impact the measured K_a values for protein-ligand interactions in aqueous solution. The most obvious factor that can induce conformational change is the change of pH (acidification or basification), thus different processes that can cause pH changes such as solvent oxidation in the ESI capillary,[100] droplet shrinkage,[101] the exposure of ESI droplet to acid or base vapors,[102-103] can all cause the protein unfolding. More interestingly, supercharging agents[104-105] enrichment by differential evaporation during droplet shrinkage, collisional and blackbody droplet heating[106] caused by the use of improper buffer like ammonium bicarbonate buffer are also shown to cause conformational changes during ESI process. As the activities happened during ESI process have become more important in the implementation of ESI-MS assays, the investigation of the protein unfolding induced by ESI process is also described in this work (Chapter 2).

1.5 The present work

The work described in this thesis focuses on the studying of ESI process and the application of direct ESI-MS assay to study noncovalent protein-ligand interactions.

The work described in chapter 2 focuses on examining ESI-induced protein unfolding. ESI-MS measurements were performed under a variety of solution conditions on a highly acidic sub-fragment (B3C) of the C-terminal carbohydrate-binding repeat region of *Clostridium difficile* toxin B, and two mutants (B4A and B4B) containing fewer acidic residues. Notably, measurements performed in negative ion mode on aqueous ammonium acetate solutions of B3C at low ionic strength (I<80 mM) revealed evidence, based on the measured charge state distribution, of protein unfolding. In contrast, no evidence of unfolding was detected from ESI-MS measurements made in positive ion mode at low I or in either mode at higher I. The results of proton nuclear magnetic resonance, taken together with the circular dichroism spectroscopy measurements and gel filtration chromatography suggest that the protein exists predominantly in a folded state in neutral aqueous solutions with I >10 mM. The results of ESI-MS measurements performed on B3C in a series of solutions with high I at pH 5 to 9 rule out the possibility that the structural changes are related to ESI-induced changes in pH. It is proposed that unfolding of B3C, observed in negative mode for solutions with low I, occurs during the ESI process and arises due to Coulombic repulsion between the negatively charged residues and liquid/droplet surface charge. ESI-MS measurements performed in negative ion mode on its mutants with less acidic residues showed support for this hypothesis.

Chapter 3 describes the deleterious non-uniform response factors induced by high molecular weight molecules and complexes in quantifying protein-ligand interactions by direct ESI-MS assay. The presence of high MW solute, that do not

interact with the protein (P) or ligand (L) of interest, is shown to result in a decrease in the abundance (*Ab*) ratio (*R*) of ligand-bound to free protein ions (i.e., *Ab*(PL)/*Ab*(P)) measured for carbohydrate-protein complexes. The magnitude of the effect is found to be more pronounced as the differences in the surface properties of P and PL become more significant. It is proposed that the decrease in *R* reflects a reduction in the number of available surface sites in the ESI droplets upon introduction of large solute and increased competition between P and the more hydrophilic PL for these available surface sites. A similar decrease in *R* that is observed upon introduction of surfactants to solution provides qualitative support for this hypothesis.

In Chapter 4, direct ESI-MS assay was used to investigate the stepwise binding of the GM1 pentasaccharide (GM1os) to the cholera toxin B subunit homopentamer (CTB_5) and to establish whether GM1os binding is cooperative. Apparent association constants were measured for the stepwise addition of one to five GM1os to CTB_5 at pH 6.9 and 22°C. The intrinsic association constant ($K_{a,int}$), which was established from the apparent association constant for the addition of a single GM1os to CTB_5, was found to be $(3.2 \pm 0.2) \times 10^6$ M^{-1} . This result is in reasonable agreement with the reported value of $(6.4 \pm 0.3) \times 10^6$ M^{-1}, which was measured using isothermal titration calorimetry at pH 7.4 and 25 °C. Analysis of the apparent association constants provides direct and unambiguous evidence that GM1os binding exhibits small positive cooperativity. Binding was found to be sensitive to the number of ligand-bound nearest neighbour subunits, with the affinities enhanced by a factor of 1.7 and 2.9 when binding occurs next to one or

two ligand-bound subunits, respectively. These findings provide quantitative support for the binding model proposed by Homans and coworkers.[107]

1.6 Literature cited

1. Ghazarian, H., B. Idoni, and S.B. Oppenheimer, *Acta Histochem.*, **2011**, 113, 236-247.

2. Dwek, R.A., *Chem. Rev.*, **1996**, 96, 683-720.

3. Neidle, S. and C.A. Bewley. 2006, Cambridge: The Royal Society of Chemistry.

4. Roldos, V., C.F. Javier, and J. Jimenez-Barbero, *ChemBioChem*, **2011**, 12, 990-1005.

5. Avci, F.Y. and D.L. Kasper, *Annu. Rev. Immunol.*, **2010**, 28, 107.

6. Varki, A., et al., Essentials of Glycobiology. 1999, New York: Cold Spring Harbor Laboratory Press. 653.

7. Ho, J.G.S., A. Greco, M. Rupnik, and K.K.S. Ng, *Proc. Natl. Acad. Sci. U. S. A.*, **2005**, 102, 18373-18378.

8. Weis, W.I. and K. Drickamer, *Annual Review of Biochemistry*, **1996**, 65, 441-473.

9. Liang, P.-H., S.-K. Wang, and C.-H. Wong, *J. Am. Chem. Soc.*, **2007**, 129, 11177-11184.

10. El-Hawiet, A., E.N. Kitova, and J.S. Klassen, *Biochem.*, **2012**, 51, 4244-4253.

11. McCammon, M.G., et al., *Structure*, **2002**, 10, 851-863.

12. Sacchettini, J.C., L.G. Baum, and C.F. Brewer, *Biochem.*, **2001**, 40, 3009-3015.

13. Bundle, D.R., *Methods Enzymol.*, **1994**, 247, 288-305.

14. Bundle, D.R., et al., *Biochemistry*, **1994**, 33, 5172.

15. Lundquist, J.J. and E.J. Toone, *Chem. Rev.*, **2002**, 102, 555-578.

16. Larsen, K., M.B. Thygesen, F. Guillaumie, W.G.T. Willats, and K.J. Jensen, *Carbohydr. Res.*, **2006**, 341, 1209-1234.

17. Meyer, B. and T. Peters, *Angew. Chem. Intl. Ed.*, **2003**, 42, 864-890.

18. Saboury, A.A., *JICS*, **2006**, 3, 1-21.

19. Wilcox, J.M., D.L. Rempel, and M.L. Gross, *Anal. Chem.*, **2008**, 80, 2365-2371.

20. Utsuno, K. and H. Uluda, *Biophys. J.*, **2010**, 99, 201-207.

21. Daghestani, H.N. and B.W. Day, *Sensors*, **2010**, 10, 9630-9646.

22. De Crescenzo, G., C. Boucher, Y. Durocher, and M. Jolicoeur, *Cell. Mol. Bioeng.*, **2008**, 1.

23. Karlsson, R., *J. Mol. Recognit.*, **2004**, 17, 151-161.

24. Schuck, P., *Annu. Rev. Biophys. Biomol. Struct.*, **1997**, 26, 541-566.

25. Homola, J., *Anal. Bioanal. Chem.*, **2003**, 377, 528-539.

26. Wishart, D., *Curr. Pharm. Biotechnol.*, **2005**, 6, 105-120.

27. Angulo, J., et al., **2006** NMR analysis of carbohydrate-protein interactions, In: Glycomics, M. Fukuda, ed.^Editors, (Elsevier Academic Press Inc), pp. 12-30.

28. Shuker, S.B., P.J. Hajduk, R.P. Meadows, and S.W. Fesik, *Science*, **1996**, 274, 1531-1534.

29. Zech, S.G., E. Olejniczak, P. Hajduk, J. Mack, and A.E. McDermot, *J. Am. Chem. Soc.*, **2004**, 126, 13948-13953.

30. Gizachew, D. and E. Dratz, *Chem. Biol. Drug Des.*, **2011**, 78, 14-24.

31. Fenn, J.B., *Angew. Chem. Int. Ed.*, **2003**, 42, 3871-3894.

32. Koichi, T., *Angew. Chem. Int. Ed.*, **2003**, 42, 3861-3870.

33. Loo, J.A., *Mass Spectrom. Rev.*, **1997**, 16, 1-23.

34. Daniel, J.M., S.D. Friess, S. Rajagopalan, S. Wendt, and R. Zenobi, *Int. J. Mass Spectrom.*, **2002**, 216, 1-27.

35. Schug, K.A., *Comb. Chem. High Throughput Screen.*, **2007**, 10, 301-316.

36. Soya, N., G.K. Shoemaker, M.M. Palcic, and J.S. Klassen, *Glycobiology*, **2009**, 19, 1224-1234.

37. Shoemaker, G.K., N. Soya, M.M. Palcic, and J.S. Klassen, *Glycobiology*, **2008**, 18, 587-592.

38. Kitova, E.N., P.I. Kitov, D.R. Bundle, and J.S. Klassen, *Glycobiology*, **2001**, 11, 605-611.

39. St. Hilaire, P.M., M.K. Boyd, and E.J. Toone, *Biochemistry*, **1994**, 33, 14452-14463.

40. Van Dongen, W.D. and A.J.R. Heck, *Analyst*, **2000**, 125, 583-589.

41. El-Hawiet, A., E.N. Kitova, L. Liu, and J.S. Klassen, *J. Am. Soc. Mass Spectrom.*, **2010**, 21, 1893-1899.

42. Wang, W., E.N. Kitova, and J.S. Klassen, *Anal. Chem.*, **2003**, 75, 4945-4955.

43. Kitova, E.N., et al., *Glycobiology*, **2007**, 17, 1127-1137.

44. El-Hawiet, A., et al., *Glycobiology*, **2011**, 21, 1217-1227.

45. Rogniaux, H., et al., *Analytical Biochemistry*, **2001**, 291, 48-61.

46. Shoemaker, G.K., N. Soya, M.M. Palcic, and J.S. Klassen, *Glycobiology*, **2008**, 18, 587-592.

47. Jecklin, M., D. Touboul, C. Bovet, A. Wortmann, and R. Zenobi, *J. Am. Soc. Mass. Spectrom.*, **2008**, 19, 332-343.

48. Donald, L.J., et al., *Protein Sci.*, **2001**, 10, 1370-1380.

49. Kempen, E.C. and J.S. Brodbelt, *Anal. Chem.*, **2000**, 72, 5411-5416.

50. Cubrilovic, D., et al., *J. Am. Soc. Mass Spectrom.*, **2012**, 23, 1768-1777.

51. El-Hawiet, A., G.K. Shoemaker, R. Daneshfar, E.N. Kitova, and J.S. Klassen, *Anal. Chem.*, **2012**, 84, 50-58.

52. El-Hawiet, A., et al., *Anal. Chem.*, **2012**, 84, 3867-3870.

53. Covey, T.R., B.A. Thomson, and B.B. Schneider, *Mass Spectrom. Rev.*, **2009**, 28, 870-897.

54. Kebarle, P. and U.H. Verkerk, *Mass Spectrom. Rev.*, **2009**, 28, 898-917.

55. Wu, X., R.D. Oleschuk, and N.M. Cann, *Analyst*, **2012**, 137, 4150-4161.

56. Rayleigh, L., *Philos. Mag.* , **1882**, 14, 184-186.

57. Cech, N.B. and C.G. Enke, *Mass Spectrom. Rev.*, **2001**, 20, 362-387.

58. Nguyen, S. and J.B. Fenn, *Proceedings of the National Academy of Sciences*, **2007**, 104, 1111-1117.

59. Konermann, L., E. Ahadi, A.D. Rodriguez, and S. Vahidi, *Anal. Chem.*, **2013**, 85, 2-9.

60. Iribarne, J.V. and B.A. Thomson, *J. Chem. Phys.*, **1976**, 64, 2287-2294.

61. Thomson, B.A. and J.V. Iribarne, *The Journal of Chemical Physics*, **1979**, 71, 4451-4463.

62. Dole, M., L.L. Mack, and R.L. Hines, *J. Chem. Phys.*, **1968**, 49, 2240-2249.

63. Iavarone, A.T. and E.R. Williams, *J. Am. Chem. Soc.*, **2003**, 125, 2319-2327.

64. Ahadi, E. and L. Konermann, *J. Phy. Chem. B*, **2012**, 116, 104-112.

65. Konermann, L., A.D. Rodriguez, and J. Liu, *Anal. Chem.*, **2012**, 84, 6798-6804.

66. Creighton, T.E., ed. *Proteins*. 1993, W. H. Freeman & Co: New York.

67. Hogan, C.J., J.A. Crarroll, H.W. Rohrs, P. Biswas, and M.L. Gross, *J. Am. Chem. Soc.*, **2008**, 130, 6926-6927.

68. Wilm, M. and M. Mann, *Anal. Chem.*, **1996**, 68, 1-8.

69. Joensson, H.N., et al., *Angew. Chem. Int. Ed.*, **2009**, 48, 2518-2521.

70. Williams, J. and K. Tomer, *J. Am. Soc. Mass Spectrom.*, **2004**, 15, 1333-1340.

71. Karas, M., U. Bahr, and T. Dulcks, *Fresenius J. Anal. Chem.*, **2000**, 366, 669-676.

72. Juraschek, R., T. Dulcks, and M. Karas, *J. Am. Soc. Mass Spectrom.*, **1999**, 10, 300-308.

73. Marshall, A.G., C.L. Hendrickson, and G.S. Jackson, *Mass Spectrom.Rev.*, **1998**, 17, 1-35.

74. Marshall, A.G. and C.L. Hendrickson, *Int. J. Mass Spectrom.*, **2002**, 215, 59-75.

75. De Hoffmann, E.S., V., ed. *Mass Spectrometry Principles and Applications*. 3 ed. 2007, John Wiley & Sons.: New York.

76. Pedder, R.E., *Ardara Technologies Technical Note*, **2009**, TN_3005A.

77. Guilhaus, M., D. Selby, and V. Mlynski, *Mass Spectrom.Rev.*, **2000**, 19, 65-107.

78. Wang, W., E.N. Kitova, and J.S. Klassen, *Methods Enzymol*, **2003**, 362, 376-397.

79. Gaylord, N.G. and J.H. Gibbs, *J. Polym. Sci.*, **1962**, 62, S22-S23.

80. Kitova, E.N., A. El-Hawiet, P.D. Schnier, and J.S. Klassen, *J. Am. Soc. Mass Spectrom.*, **2012**, 23, 431-441.

81. Yanes, O., J. Villanueva, E. Querol, and F.X. Aviles, *Nat. Protoc.*, **2007**, 2, 119-130.

82. Jecklin, M.C., et al., *Anal. Chem.*, **2009**, 81, 408.

83. Yanes, O., J. Villanueva, E. Querol, and F.X. Aviles, *Mol. Cell Proteomics*, **2005**, 4, 1602.

84. Gross, J., A. Xavier, and M.L. Gross. *Abstracts of Papers*. in *49th ASMS Conference on Mass Spectrometry and Allied Topics*. 2001. Chicago, IL.

85. Chitta, R.K., D.L. Rempel, and M.L. Gross, *J. Am. Soc. Mass Spectrom*, **2005**, 16, 1031-1038.

86. Gabelica, V., N. Galic, F. Rosu, C. Houssier, and E. De Pauw, *J. Mass Spectrom.*, **2003**, 38, 491-501.

87. Mathur, S., S. Badertscher, M. Scott, and R. Zenobi, *Phys. Chem. Chem. Phys.*, **2007**, 9, 6187-6198.

88. Wortmann, A., F. Rossi, G. Lelais, and R. Zenobi, *J. Mass Spectrom.*, **2005**, 40, 777-784.

89. Sun, J., E.N. Kitova, W. Wang, and J.S. Klassen, *Anal. Chem.*, **2006**, 78, 3010-3018.

90. Hossain, B.M. and L. Konermann, *Anal. Chem.*, **2006**, 78, 1613-1619.

91. Sun, J., E.N. Kitova, N. Sun, and J.S. Klassen, *Anal. Chem.*, **2007**, 79, 8301-8311.

92. Sun, N., J. Sun, E.N. Kitova, and J.S. Klassen, *J. Am. Soc. Mass Spectrom*, **2009**, 20, 1242-1250.

93. Sun, N., N. Soya, E.N. Kitova, and J.S. Klassen, *J. Am. Soc. Mass Spectrom.*, **2010**, 21, 472-481.

94. Kitova, E.N., N. Soya, and J.S. Klassen, *Anal. Chem.*, **2011**, 83, 5160-5167.

95. Deng, L., N. Sun, E.N. Kitova, and J.S. Klassen, *Anal. Chem.*, **2010**, 82, 2170-2174.

96. Liu, L., D. Bagal, E.N. Kitova, P.D. Schnier, and J.S. Klassen, *J. Am. Chem. Soc.*, **2009**, 131, 15980-15981.

97. Hunter, E.P.L. and S.G. Lias, *J. Phys. Chem. Ref. Data*, **1998**, 27, 413-656.

98. Bagal, D., et al., *Anal. Chem.*, **2009**, 81, 7801-7806.

99. Liu, L., E.N. Kitova, and J.S. Klassen, *J. Am. Soc. Mass Spectrom.*, **2011**, 22, 310-318.

100. Konermann, L., E.A. Silva, and O.F. Sogbein, *Anal. Chem.*, **2001**, 73, 4836-4844.

101. Zhou, S., B.S. Prebyl, and K.D. Cook, *Anal. Chem.*, **2002**, 74, 4885-4888.

102. Kharlamova, A., B.M. Prentice, T.-Y. Huang, and S.A. McLuckey, *Anal. Chem.*, **2010**, 82, 7422-7429.

103. Kharlamova, A. and S.A. McLuckey, *Anal. Chem.*, **2011**, 83, 431-437.

104. Lomeli, S., I. Peng, S. Yin, R. Ogorzalek Loo, and J. Loo, *J. Am. Soc. Mass Spectrom.*, **2010**, 21, 127-131.

105. Sterling, H., et al., *J. Am. Soc. Mass Spectrom.*, **2012**, 23, 191-200.

106. Sterling, H.J., C.A. Cassou, A.C. Susa, and E.R. Williams, *Anal. Chem.*, **2012**, 84, 3795-3801.

107. Turnbull, W.B., B.L. Precious, and S.W. Homans, *J. Am. Chem. Soc.*, **2004**, 126, 1047-1054.

Chapter 2

Electrospray Ionization-Induced Protein Unfolding[*+]

2.1 Introduction

Electrospray ionization mass spectrometry (ESI-MS) has become an important research tool in the fields of structural and chemical biology.[1-8] It is used to probe the higher-order structure and conformation of proteins and protein assemblies,[4, 6] to investigate protein folding/unfolding pathways,[9-11] to identify noncovalent protein interactions and to quantify the corresponding kinetic and thermodynamic parameters.[7, 12-16] The various experimental strategies available to characterize the higher-order structures of proteins and their complexes can be classified as either "direct" or "indirect" in their approach. Indirect methods typically involve ESI-MS analysis of proteolytic peptides to investigate protein conformation and interactions. For example, the extent of peptide backbone amide hydrogen-deuterium exchange can be used to interrogate the structure and dynamics of proteins and protein complexes.[7, 11, 17] Related and complementary approaches, based on oxidative labeling[18-19] or chemical cross-linkers,[20] are also employed to deduce structural information.[21] Direct methods rely on the analysis of the intact protein or protein complex in the gas phase by ESI-MS alone or in combination with other gas phase techniques. Although the factors responsible for protein ionization in ESI are not fully understood, it is generally agreed that the size of

[*] A version of this chapter has been published: Lin, H.; Kitova, E. N.; Johnson, M. A.; Eugenio, L; Ng, K. K. S.; Klassen, J. S. *J. Am. Soc. Mass Spectrom.* 2012, 23, 2122-2131.
[+] Protein expression and purification were done by Eugenio, L. (University of Calgary).

the protein (in particular the solvent accessible surface area) is a major determining factor.[22-23] Consequently, the measured charge state distribution (CSD) can be used to establish the presence of folded and partially or fully denatured proteins in solution and to monitor conformational changes in response to changes in solution conditions.[9, 24-29] Collision cross sections (CCS) of the gaseous ions of proteins and protein complexes, as determined by ion mobility spectrometry, can also provide insight into protein structure in solution.[30-32]

That the ESI process itself does not significantly perturb protein structure represents an important underlying assumption in the implementation of direct ESI-MS methods to characterize the structure of proteins and protein complexes in solution or structural changes resulting from changes to solution conditions. There exist abundant data to suggest that this is generally true. For example, the detection of noncovalent protein interactions, *e.g.*, protein-ligand and protein-protein complexes,[1-7, 13, 15-16, 33-35] by ESI-MS would not be possible if proteins were to undergo significant structural changes during the ESI process. In fact, there is growing evidence that the solution specific intermolecular interactions in protein complexes are, to some extent, preserved in the gas phase.[30, 33-35] The similarities found between the CCS values measured for many gaseous protein ions and those estimated from their crystal structures also argue against widespread protein structural changes during ESI.[30, 36-37] The general preservation of the native structure or close-to-native structure of proteins and protein complexes in the gas phase can be rationalized based on the timescale of the ESI process, which is typically in the μs - ms range.[38-40] Consequently, only relatively

fast structural rearrangement processes will have a chance to occur during the life of the ESI droplets.[41-42]

While it is generally safe to conclude, based on the abundant data described above, that protein structures and intermolecular interactions are not dramatically perturbed by the ESI process, lately evidence has emerged to show that this is not true in all situations. McLuckey and coworkers demonstrated that the charge states of gaseous protein ions formed by ESI from aqueous solutions can be significantly increased by introducing acidic or basic vapour into the ESI source.[43-45] These observations were attributed to pH-induced changes resulting from the dissolution of the acidic or basic vapour into the ESI droplets. The addition of low-volatility reagents, such as *m*-nitrobenzyl alcohol, dimethyl sulfoxide (DMSO) and sulfolane, to aqueous protein solutions has been shown in some cases to produce a significant increase in the protein ion charge states, i.e., supercharging.[41, 46-48] Williams and coworkers have attributed the increased charging to the rapid increase in reagent concentration in the ESI droplets due to solvent evaporation, which promotes thermal or chemical denaturation of the protein.[48] The hypothesis that a rapid increase in reagent concentration in the droplets promotes protein unfolding is supported by the results of a recent study by Julian and coworkers on the unfolding of myoglobin in a non-denaturing solution of water and DMSO resulting from partial lyophilization of the sample.[26] Recently, Williams and coworkers demonstrated the formation of highly charged protein ions from aqueous ammonium bicarbonate solutions at neutral pH by ESI performed in positive ion mode.[49] The enhanced charging, which was found to be

sensitive to many factors including the temperature of the entrance capillary of the mass spectrometer, the spray potential and the ionic strength of the solution, was attributed to thermal denaturation of the protein resulting from the rapid heating of the ESI droplets in the atmosphere-vacuum interface.[49] Interestingly though, measurements performed on solutions containing ammonium acetate (NH_4OAc) yielded no evidence of protein unfolding.

Here, we report new evidence for protein unfolding induced by the electrospray ionization process. A series of ESI-MS measurements were performed on a highly acidic sub-fragment (B3C) of the C-terminal carbohydrate-binding repeat region of the large exotoxin, toxin B (TcdB), produced by *Clostridium difficile*[50-51] and two mutants engineered to replace negatively charged residues with Ala residues, under a variety of solution conditions. The ESI-MS measurements performed on aqueous NH_4OAc solutions at low ionic strength (I <80 mM) revealed evidence, based on the measured charge state distribution, of protein unfolding in negative ion mode, but not in positive ion mode. In contrast, no evidence of unfolding was found from the ESI-MS measurements performed in either mode at high I. The results of proton nuclear magnetic resonance (NMR) and circular dichroism (CD) spectroscopy measurements, as well as gel filtration chromatography (GFC), performed on B3C under low and high I conditions indicate that the protein exists predominantly in a folded state in neutral aqueous solutions and that the structure is not strongly dependent on solution I. The results of control experiments confirmed that the structural changes are not related to ESI-induced changes in solution pH. It is

proposed that the unfolding of B3C, which is observed in negative ion mode for solutions with low *I*, occurs in the ESI droplets and is electrostatically-driven. ESI-MS measurements performed in negative ion mode on mutants B4A and B4B, which contain fewer acidic residues than B3C, also reveal a shift to higher (negative) charge states at low *I*. However, in both cases, the magnitude of the change is smaller than observed for B3C.

2.2 Experimental section

2.2.1 Proteins

The TcdB-B3C fragment consists of an N-terminal Met residue, followed by six His residues and 254 residues from the carboxy-terminus of the toxin TcdB from *Clostridium difficile* strain 630.[52] In addition to the seven non-natural residues added to the N-terminus, two charged residues at positions 142 and 143 of B3C (Glu2246 and Lys2247 in wild-type TcdB) were replaced with Ala as a part of an unrelated study attempting to improve the crystallization properties of the protein sub-fragment. This sub-fragment from the carboxy-terminal carbohydrate-binding repeat region of TcdB contains a large excess of negatively charged residues (48 acidic Glu and Asp residues versus 13 basic Arg and Lys residues), which is characteristic of the entire repeat region.

Two artificial gene variants were also synthesized (Genscript) and cloned into the pGS-21a expression plasmid to produce two protein fragments with multiple negatively charged residues replaced by the neutral residue Ala. In B4A, Ala replaces Asp at positions 222, 224 and 244, as well as replacing Glu at

positions 225, 246 and 247. B4B is identical to TcdB-B4A, except that Ala also replaces Asp at positions 109, 134 and 135, as well as replacing Glu at positions 111 and 113. However, the two charged residues (Glu at position 142 and Lys at position 143) mentioned above have been retained in the mutants rather than being replaced by Ala. As a result, B4A contains, in total, 43 acidic residues (5 fewer than B3C) while B4B contains, in total, 38 acidic residues (10 fewer than B3C). The calculated isoelectric points (*pI*) of B3C, B4A and B4B (http://web.expasy.org/protparam/) are 4.08, 4.21 and 4.32, respectively.

Each protein was expressed from a synthetic gene (Genscript) containing codons optimized for high-level expression in *Escherichia coli*. The genes were expressed from the T7 promoter in pGS-21a (Genscript) using *E. coli* C41 (DE3) as a host. Cells were grown in LB autoinduction medium (Formedium) at 26 °C for 24 h before harvesting by centrifugation, resuspension in (50 mM sodium phosphate, pH 8.0, 10 mM imidazole chloride, 300 mM sodium chloride, 50 g/L glycerol) and stored at -80°C. The proteins were purified as previously described [52] with the addition of a final gel filtration chromatography (Sephacryl S-300 HR) step using buffer B (25 mM Tris-Cl pH 7.5, 150 mM sodium chloride, 0.5 mM EDTA, 5% (w/v) glycerol).

For the solution NMR measurements, B3C was dialyzed using Spectra/Por 4 (12-14 kDa molecular weight cut-off (MWCO)) dialysis membrane (Spectrum) for 18 h against 9.8 mM sodium/potassium phosphate buffer (pH 7.0) containing either 5 mM NaCl (low *I* sample) or 150 mM NaCl (high *I* sample). For the high *I* sample, 560 μL of protein solution was obtained and 70 μL of phosphate-buffered

saline (10 mM sodium phosphate, pH 7.2, 137 mM NaCl) and 50 μL of D_2O were added to yield 680 μL of 0.3 mM B3C. For the low *I* sample, 50 μL of D_2O was added to 650 μL of protein solution to yield 700 μL of sample at 0.3 mM B3C. For the CD measurements, a series of solutions of 15 μM B3C were prepared with varying concentrations of sodium/potassium phosphate buffer to give *I* values of 20 mM, 60 mM and 105 mM. As a control, a 15 μM B3C solution containing 60 mM of sodium/potassium phosphate buffer and 6 M guanidinium chloride was also prepared and used to obtain the CD spectrum of unfolded (denatured) B3C. For the gel filtration chromatography, samples of B3C (20 μM) were dialyzed for 16 hours against each of the solutions used for gel filtration (10 mM sodium/potassium phosphate (pH 7.0) and containing 5 or 150 mM NaCl) immediately prior to injection on the column. For the ESI-MS experiments, B3C, B4A and B4B were each concentrated and exchanged with aqueous 50 mM ammonium acetate (pH 7) using ultracentrifugation microconcentrators (Millipore Corp., Bedford, MA) with a 10 kDa MWCO and stored at –20 ^{o}C if not used immediately. For all ESI-MS measurements, the protein concentration was fixed at 15 μM. Ammonium acetate was added to the samples to yield different final concentrations ranging from 10 mM to 200 mM. The corresponding *I* values are similar in magnitude, ranging from 9.6 mM to 198.1 mM. For simplicity, the solution *I* values reported for the ESI-MS measurements are taken to be equal to ammonium acetate concentrations. Acetic acid was added for low pH solutions and ammonium hydroxide was added for high pH solutions.

2.2.2 Mass spectrometry measurements

In all cases, ESI was performed using nanoESI tips pulled from borosilicate glass capillaries (1.0 mm o.d., 0.78 mm i.d.) using a P-97 micropipette puller (Sutter Instruments, Novato, CA). ESI-MS measurements were performed on a Synapt G2 quadrupole-ion mobility separation-time-of-flight (Q-IMS-TOF) mass spectrometer (Waters UK Ltd., Manchester, UK) and an ApexII 9.4 tesla Fourier transform ion cyclotron resonance (FTICR) mass spectrometer (Bruker, Billerica, MA).

ApexII 9.4T FTICR mass spectrometer. Details of the standard instrumental and experimental conditions used for ESI-MS analysis of proteins with this instrument are described elsewhere.[16]

Synapt G2 Q-IMS-TOF mass spectrometer. Mass spectra were obtained in either positive or negative ion modes using cesium iodide (concentration 30 ng μL^{-1}) for calibration. Given below are some of the instrumental conditions used to carry out the measurements in positive ion mode. A capillary voltage of 1.4 kV under positive mode was applied to carry out nanoESI. A cone voltage of 40 V was used and the source block temperature was maintained at 70 ºC. Other important voltages for ion transmission, that is the injection voltages into the trap and transfer ion guides, were maintained at 10 V and 5 V, respectively. Argon was used in the trap and transfer ion guides at a pressure of 2.22×10^{-2} mbar and 3.36×10^{-2} mbar, respectively. Data acquisition and processing were carried out using MassLynx (v 4.1).

2.2.3 Average Charge State Calculation

The average charge state (*ACS*) of the protein ions was calculated from the ESI mass spectrum using eq 1:

$$ACS = \frac{\sum_n I_n n}{\sum_n I_n} \quad (1)$$

where I_n is the protein ion intensity (measured as peak height) and n is the charge state.

2.2.4 NMR measurements

1D ^{1}H NMR spectra were recorded at 21 °C on a Varian VNMRS 700 MHz spectrometer equipped with a 5mm ^{1}H{^{13}C/^{15}N} z-gradient cryogenic probe. A total of 1024 transients, consisting of 19685 complex points spanning a spectral width of 9842 Hz, were collected. The data were zero-filled to 32768 complex points and multiplied by an exponential apodization function with broadening constant of 1 Hz before Fourier transformation. The H_2O signal was suppressed using WATERGATE.[53]

2.2.5 Circular dichroism measurements

Circular dichroism (CD) spectra were recorded at 20.8 °C on an OLIS DSM CARY-17 spectrophotometer conversion and circular dichroism module (On-line Instrument Systems Inc.) using a 0.2 mm path length quartz cuvette. Data were collected in scanning mode from 300 nm to 190 nm and the average value of 5

repetitions was reported. Data were analyzed with OLIS Spectral Works (Version 4.3), converting into molar ellipticity units. For each buffer condition, the spectrum of the CD buffer alone was subtracted from the spectrum of the sample containing protein.

2.2.6 Gel filtration

Gel filtration chromatography (GFC) was performed using a Superose 6 Tricorn Column (GE Healthcare, 10 mm ID X 300 mm, 24 mL bed volume) equilibrated in 10 mM sodium/potassium phosphate (pH 7.0) and containing 5 or 150 mM NaCl. The running buffer was degassed under vacuum immediately before connecting to the chromatographic system and the column was run under a constant flow rate of 0.4 mL/min (~200 psi total system pressure) using a Shimadzu Prominence HPLC system with LC-20AD pumps and a SPD-20AV detector measuring absorbance at 280 nm. Samples of B3C (20 μM, 20 μL) were loaded onto the column using a manual injector (Rheodyne). Elution volumes and peak areas were evaluated using CLASS-VP (Shimadzu) software.

2.3 Results and Discussion

ESI mass spectra were measured in both positive and negative ion mode for aqueous solutions of B3C (15 μM) and NH_4OAc at concentrations ranging from 10 mM to 200 mM (pH 7). Shown in Figure 2.1 are representative mass spectra acquired in both modes for solutions containing 10 mM, 80 mM and 200 mM NH_4OAc. In all cases, the major protein ions detected correspond to multiply

protonated or deprotonated B3C, i.e., $(B3C+nH)^{n+} \equiv P^{n+}$ (positive ion mode), $(B3C-nH)^{n-} \equiv P^{n-}$ (negative ion mode). Notably, the CSD observed in positive ion mode is found to be relatively insensitive to the concentration of NH_4OAc, with n ranging from +9 to +11 and an *ACS* of +10.0 ± 0.1 (Figure 2.2). At NH_4OAc concentrations above ~80 mM, the CSD measured in negative ion mode is similar (in terms of the number of charges) to that observed in positive ion mode, with an *ACS* of -10.2 ± 0.1 (Figures 2.1c and 2.1d; Figure 2.2). However, lower NH_4OAc concentrations produce P^{n-} ions with a much broader CSD. For example at 10 mM NH_4OAc, the P^{n-} ions charge states range from -10 to -23; with a corresponding *ACS* of -16.6. The aforementioned results were acquired using the Waters Synapt G2 Q-IMS-TOF mass spectrometer. However, the present findings are independent of the instrumentation used to collect the ESI mass spectra, with similar results obtained using a 9.4T FTICR instrument (Figure 2.3).

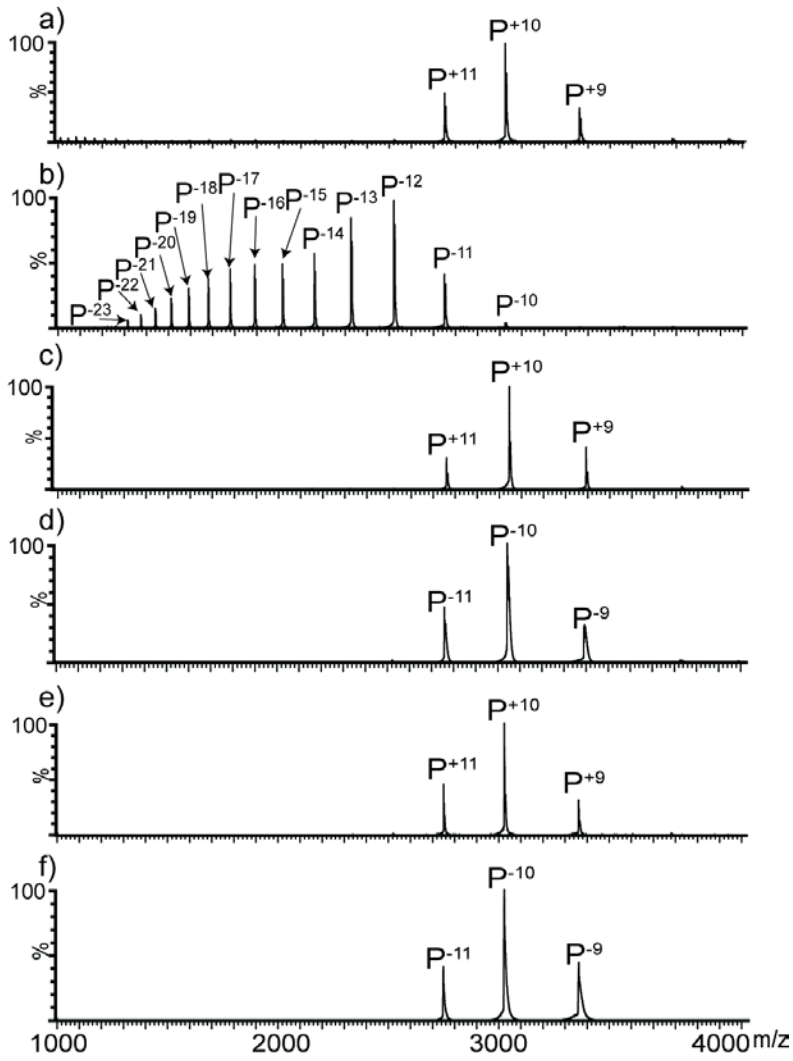


Figure 2.1 ESI mass spectra acquired in positive ion mode for aqueous solutions of TcdB-B3C (P, 15 μM) and ammonium acetate (a) 10 mM, (c) 80 mM and (e) 200 mM. ESI mass spectra acquired in negative ion mode for aqueous solutions of TcdB-B3C (P, 15 μM) and ammonium acetate (b) 10 mM, (d) 80 mM and (f) 200 mM.

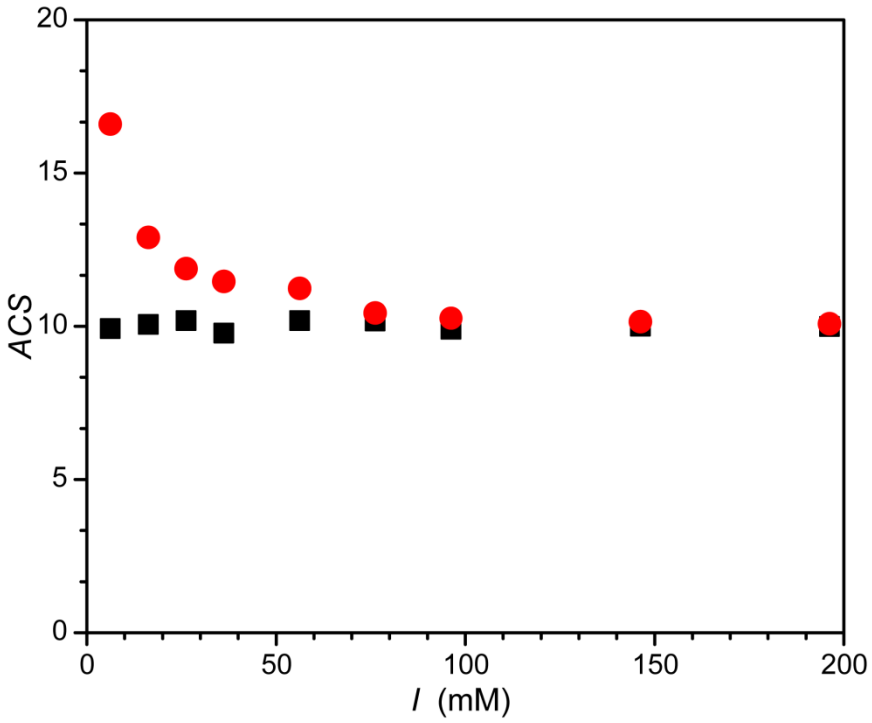


Figure 2.2 Plot of average charge state (*ACS*) versus ionic strength (*I*) measured from ESI mass spectra acquired in positive (■) and negative (●) ion modes for aqueous ammonium acetate solutions (pH 7) of TcdB-B3C (15 μM).

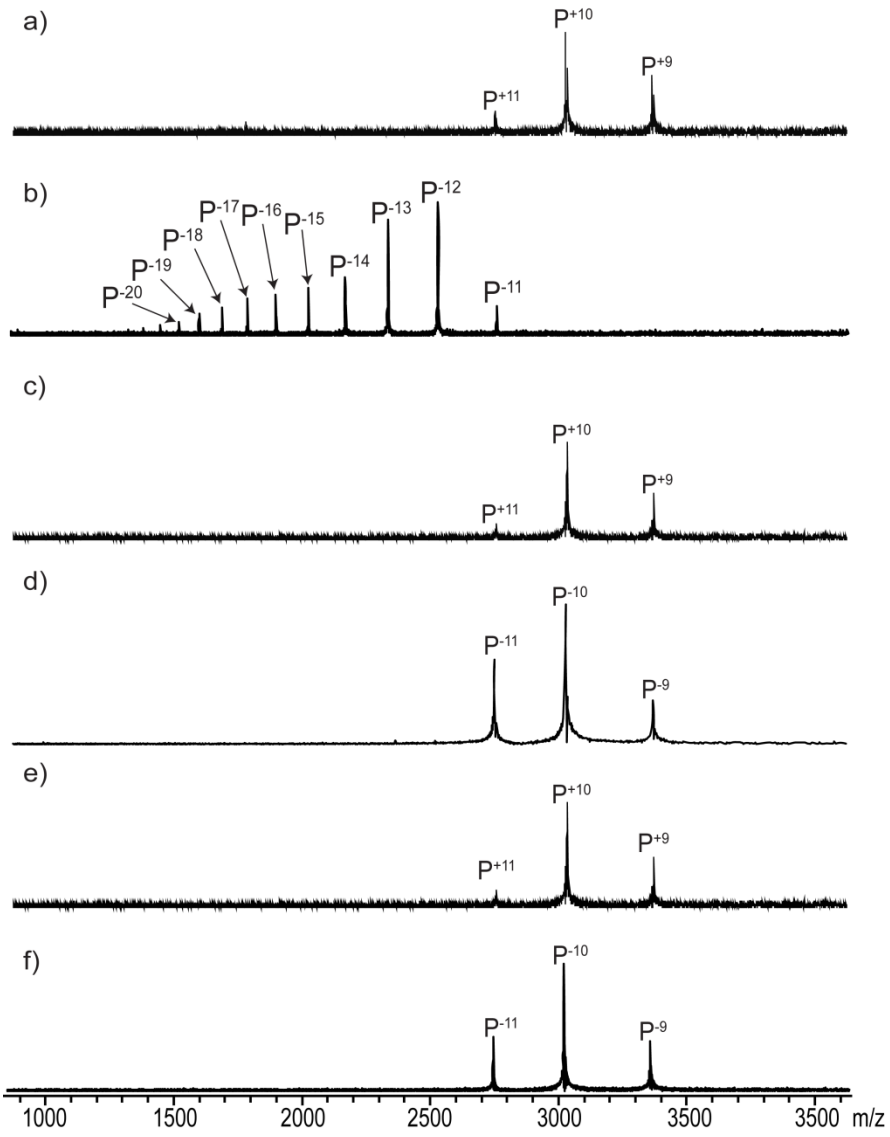


Figure 2.3 ESI mass spectra acquired in positive mode for aqueous solutions of TcdB-B3C (P, 15 μM) and ammonium acetate (a) 10 mM; average charge state (*ACS*) +9.58, (c) 80 mM; *ACS* +9.81 and (e) 200 mM; *ACS* of +9.91. ESI mass spectra acquired in negative mode for aqueous solutions of TcdB-B3C (P, 15 μM) and ammonium acetate (b) 10 mM; *ACS* -15.11, (d) 80 mM; *ACS* -9.93 and (f) 200 mM; *ACS* -10.01. All measurements carried out using a Bruker Apex II 9.4T FTICR MS.

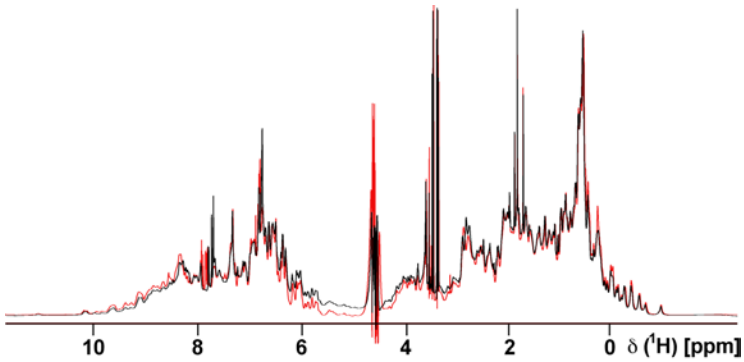
The narrow CSD observed in both modes for the solutions at higher *I* suggest that B3C exists in a compact, folded form in neutral aqueous solution. However, the broadening of the CSD and shift to higher charge states observed in negative ion mode for the solutions at low *I* (< 80 mM) suggests that B3C is at least partially unfolded under these conditions. The observation of protein unfolding at low buffer concentrations (i.e., low *I*) is not, in itself, remarkable. However, it is intriguing that the ESI mass spectra acquired for the same solution, but in positive ion mode, show no evidence (based on the measured CSD) of protein unfolding. There are a number of possible explanations for these seemingly contradictory observations. It is possible that both folded and unfolded B3C co-exist in solution (at least at low *I*) and that, because of differences in relative response factors [54], ions corresponding to the unfolded protein are more abundant than those of the folded structure in negative ion mode than in positive ion mode. It also conceivable that, as reported by Kaltashov and coworkers,[29] asymmetric dissociation of protein aggregates in the gas-phase alters the charge state distribution of protein ions, giving the appearance of protein unfolding in solution.[55] However, given that gentle sampling conditions (suitable for detecting protein-ligand complexes) were employed, this explanation seems unlikely. It is also possible that unfolding occurs selectively in negative ion mode in response to an increase in the local pH of the solution at the end of the ESI tip due to electrochemical reduction of the solvent.[56-57] An alternative possibility is that the ESI process itself induces protein unfolding in the droplets, at least in negative ion mode. While it is difficult to devise experiments to directly probe protein structure

or changes in structure in the droplets produced by nanoESI, it is, in principle, possible to establish the presence of different forms of B3C in bulk solution or structural changes resulting from changes in solution composition (e.g. *I*, pH).

Several lines of investigation were undertaken to establish whether B3C exists, at least in part, in an unfolded form in neutral aqueous solution at low *I*. One-dimensional (1D) ^{1}H NMR analysis was used to analyze the structure of B3C in neutral aqueous solution with low (20 mM) and high *I* (190 mM). The physical basis of the relationships between 1D NMR observables, such as chemical shift dispersion and line-width, and protein tertiary and quaternary structure are well understood[58] and are commonly used to distinguish globular (folded) proteins from partly or completely denatured proteins, natively unfolded proteins, and other intermediate folding states.[59-62] As seen in Figure 2.4a, the primary characteristic indicating globular folding for B3C is the high level of ^{1}H chemical shift dispersion throughout the entire spectrum, including the regions corresponding to methyl protons (-1 to 1.5 ppm), α-protons (3.5 to 6 ppm) and amide protons (6 to 10 ppm). The chemical shift dispersion in these areas results from the variety of local microenvironments created by the three-dimensional protein structure, over and above what is expected from residual structure present in unfolded polypeptide chains.[63-66] That the NMR spectra are similar under both high and low *I* conditions suggests that the structure of B3C is not strongly influenced by *I* and that it exists predominantly in a globular form. However, due to sensitivity and signal overlap considerations, it is not possible to rule out the presence of a small fraction of unfolded protein.

The influence of *I* on the structure of B3C was also investigated by CD spectroscopy. Spectra were recorded for B3C solutions with *I* ranging from 20 mM to 105 mM (Figure 2.4b). Although the spectra do exhibit subtle differences over this range of *I*, the results are not consistent with a significant change in secondary structure. Furthermore, upon addition of 6 M guanidium chloride, a strong denaturant, a dramatic change in the CD spectrum is evident, consistent with the loss of ordered structure and suggesting that the majority of protein in solution is folded.

a)



b)

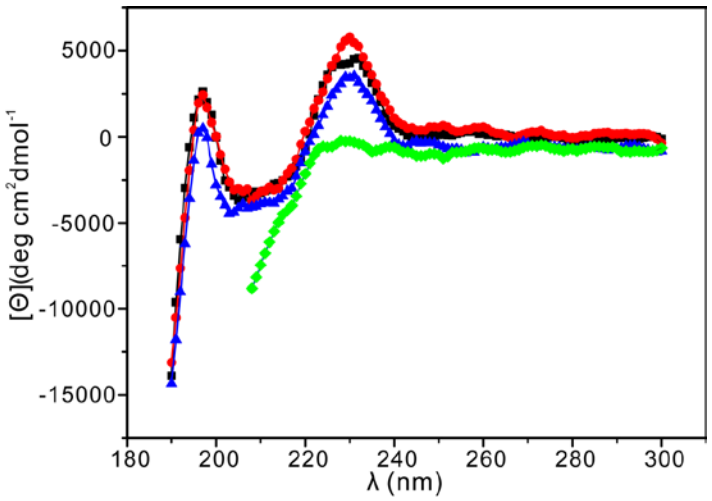
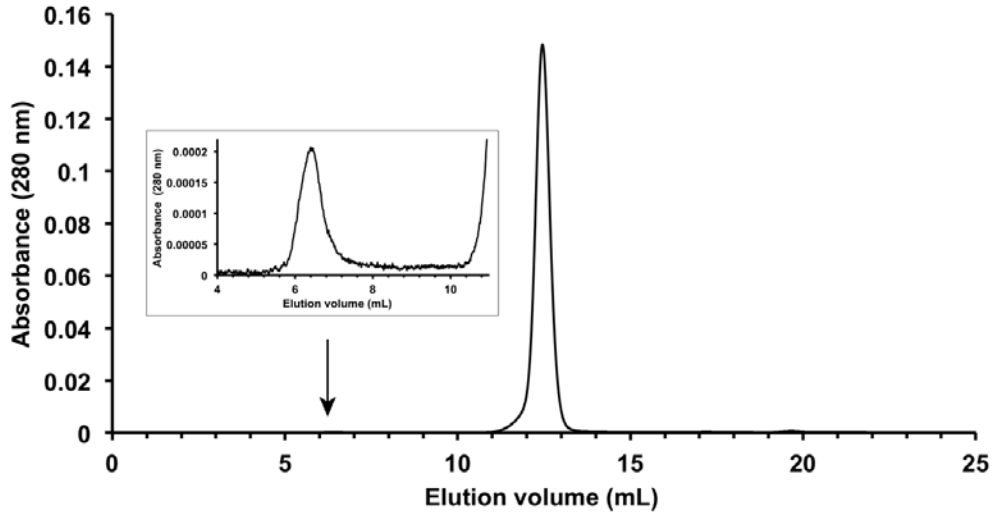


Figure 2.4 (a) 1D ^{1}H NMR spectra for neutral aqueous solutions of TcdB-B3C. The spectrum shown in black was measured with 0.3 mM B3C in phosphate buffer (pH 7.0) with 150 mM NaCl. The spectrum shown in red was measured with 0.3 mM B3C in phosphate buffer (pH 7.0) with 5 mM NaCl. The spectra were recorded at 21 °C, with 1024 scans and with WATERGATE water suppression,[53] on a Varian VNMRS 700 MHz spectrometer equipped with a 5mm ^{1}H{^{13}C/^{15}N} z-gradient cryogenic probe. The sharp lines between 3.2 and 3.6 ppm derive from the concentrator membrane. (b) CD spectra of aqueous solutions of TcdB-B3C with phosphate buffer at different ionic strengths: 20 mM, ■; 60 mM, ●; 105 mM, ▲; and with the addition of 6 M guanidinium chloride, ♦.

To quantitatively probe for a population of unfolded protein in solution, GFC was performed on B3C solutions with both high (190 mM) and low *I* (20 mM). Under both conditions B3C elutes almost completely in a single Gaussian-shaped peak (Figure 2.5). The elution volume was earlier than expected for a protein of this size, and this effect was magnified at low *I*. This effect has previously been observed in polymers bearing the same net charge as the matrix. Electrostatic repulsion between the polymer and the matrix leads to an ionic exclusion effect that reduces the effective volume available to the charged polymer, thus decreasing elution volumes relative to uncharged polymers of the same size.[67] Despite this effect, the elution volume of B3C under conditions of low or high *I* is roughly double that of the void volume, where aggregated protein with non-native three-dimensional structure is expected. Also, the lack of other significant peaks in the chromatogram strongly indicates that unfolded forms of B3C are not present in significant amounts. Assuming that the void volume peak, which is the only significant additional peak present in the elution profile, contains unfolded protein, the total fraction of unfolded protein in the sample at either low *I* or high *I* is estimated to be <0.1%.

a)



b)

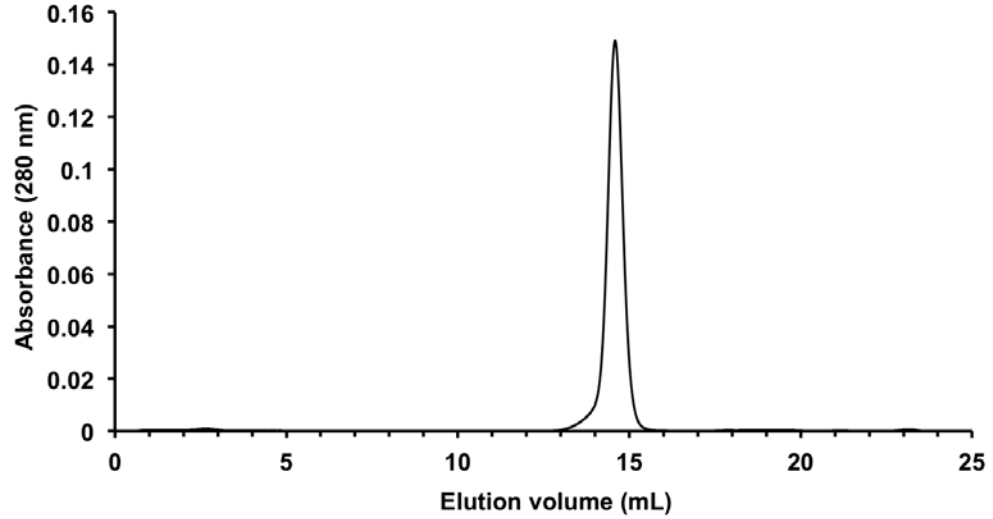


Figure 2.5 Chromatograms of B3C (12 μg) eluted from Superose 6 10/300 gel filtration column equilibrated with 10 mM sodium/potassium phosphate (pH 7.0) and (a) 5 mM sodium chloride or (b) 150 mM sodium chloride. The inset in panel (a) shows a magnified view of the void volume peak.

Taken together, the results of the NMR and CD spectroscopy and GFC measurements suggest that B3C exists predominantly in a folded state in neutral aqueous solutions with *I* values of 20 to 190 mM. It follows then, that the differences in CSD observed for B3C in negative ion mode ESI-MS under low and high *I* conditions do not reflect protein structural changes in bulk solution.

If the differences in CSD observed in positive and negative ion mode for the solutions with low *I* do not reflect the presence of both folded and unfolded B3C in bulk solution (and differences in relative response factors), then the differences must be due to the ESI process itself. To establish whether the unfolding of B3C occurs selectively in negative ion mode (for solutions with low *I*) due to an electrochemically-induced increase in the local pH at the end of the ESI tip, ESI-MS measurements were performed on aqueous solutions of B3C with high *I* (200 mM) at pH values ranging from 5 to 9 in both positive and negative ion modes (Figure 2.6). Although there are subtle differences in the appearance of the mass spectra measured for the solutions at different pH values, in particular the extent of adduct formation observed in negative ion mode, the CSD of the B3C ions exhibits no significant dependence on the solution pH in either mode. These results conclusively rule out the possibility that changes in pH are responsible for the differences in CSD observed in negative ion mode for solutions of B3C at low and high *I*.

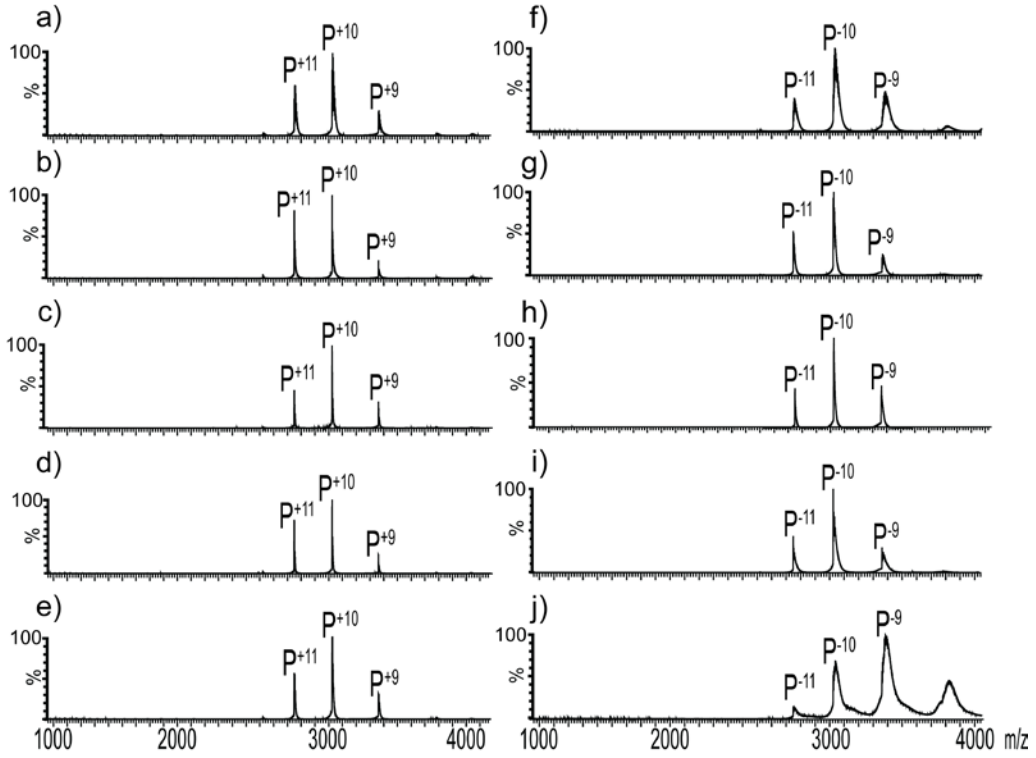


Figure 2.6 ESI mass spectra acquired in positive ion mode for aqueous solutions of TcdB-B3C (P, 15 μM) and ammonium acetate (200 mM) at pH (a) 9.0, (b) 8.0, (c) 7.0, (d) 6.0 and (e) 5.0. ESI mass spectra acquired in negative ion mode for aqueous solutions of TcdB-B3C (P, 15 μM) and ammonium acetate (200 mM) at pH (f) 9.0, (g) 8.0, (h) 7.0, (i) 6.0 and (j) 5.0.

If pH changes are not responsible for the difference in the CSD measured for B3C at low and high *I* in negative ion mode then one is forced to consider the possibility that unfolding occurs selectively during the ESI process carried out in negative ion mode, albeit only for solutions at low *I*. But what is the driving force for unfolding? B3C is an unusually acidic protein and is expected to have high surface activity owing to the large excess of negative charge present at neutral pH. Consequently, it is conceivable that Coulombic repulsion between the negatively

charged residues and the ESI droplet surface charge induces unfolding. That unfolding is not observed in negative ion mode for solutions with higher *I* can be attributed to the effective shielding of the negative charges by the buffer ions. It has been shown in a number of studies that proteins in aqueous solution can be denatured by relatively high electric fields.[68-70] Therefore, it is also possible that unfolding occurs at the ESI tip prior to droplet formation.

To test the hypothesis that B3C undergoes charge-induced unfolding in negative ion mode, ESI-MS measurements were performed on neutral aqueous solutions of two mutant proteins, B4A and B4B, which have fewer (5 and 10, respectively) acidic residues than B3C, and NH_4OAc concentrations ranging from 10 to 200 mM. Illustrative mass spectra acquired in both modes for solutions containing 15 μM mutant protein and NH_4OAc at three different concentrations are shown for B4A and B4B in Figure 2.7 and 2.8, respectively. The corresponding plots of *ACS* versus *I* are shown in Figure 2.9. It can be seen that for both mutants the mass spectra acquired for solutions with low and high *I* are qualitatively similar to those measured for B3C (Figure 2.1). For solutions with high *I*, the *ACS* measured for B4A and B4B in positive and negative ion modes (10.6 ± 0.1 and -10.6 ± 0.1, and 10.6 ± 0.1 and -10.3 ± 0.3, respectively) are similar to the results obtained for B3C. For the solutions at low *I* (10 mM) there is no evidence of unfolding in positive ion mode. In contrast, in negative ion mode the *ACS* increases significantly, to -14.0 for B4A and -12.7 for B4B, which is consistent with unfolding. However, the absolute *ACS* values determined under these conditions are noticeably smaller than those measured for B3C. Furthermore,

inspection of the plot of *ACS* versus *I* reveals that the onset of unfolding for the two mutants occurs at lower NH_4OAc concentrations than found for B3C. For B4A, the *ACS* values start to increase (become more negative) at concentrations <60 mM, while for B4B the change occurs at concentrations <40 mM. These differences are more clearly seen in the mass spectra measured for B3C, B4A and B4B at NH_4OAc concentrations of 20 mM, 40 mM and 60 mM (Figures 2.10, 2.11 and 2.12). That the increase in *ACS* in negative ion mode occurs at lower *I* is consistent with the reduction in the electrostatic repulsion (the putative driving force for protein unfolding) between the mutant proteins, which contain fewer acidic residues, and the droplet surface charge.

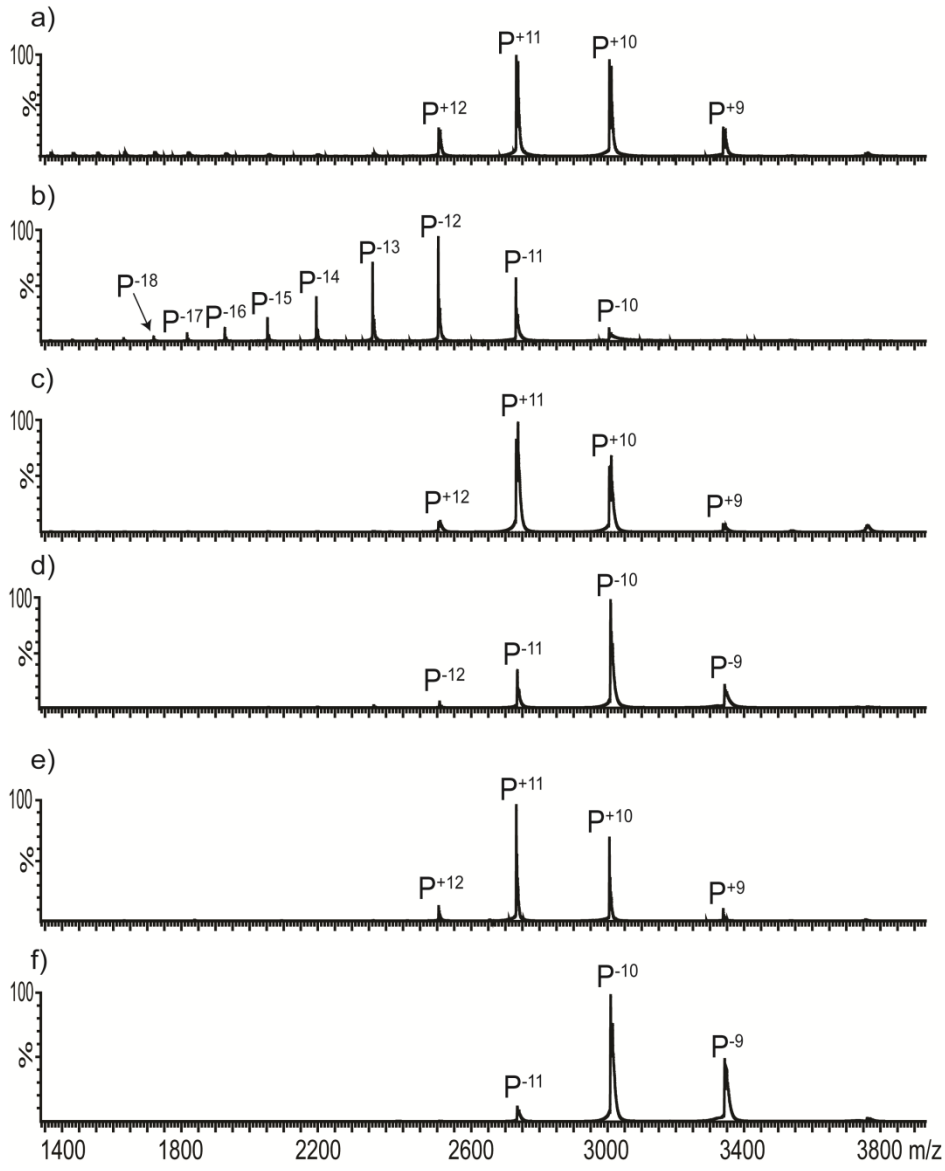


Figure 2.7 ESI mass spectra acquired for aqueous solutions of TcdB-B4A (P, 15 μM) with 10 mM ammonium acetate in (a) positive mode and (b) negative mode; 60 mM ammonium acetate in (c) positive mode and (d) negative mode; 200 mM ammonium acetate in (e) positive mode and (f) negative mode.

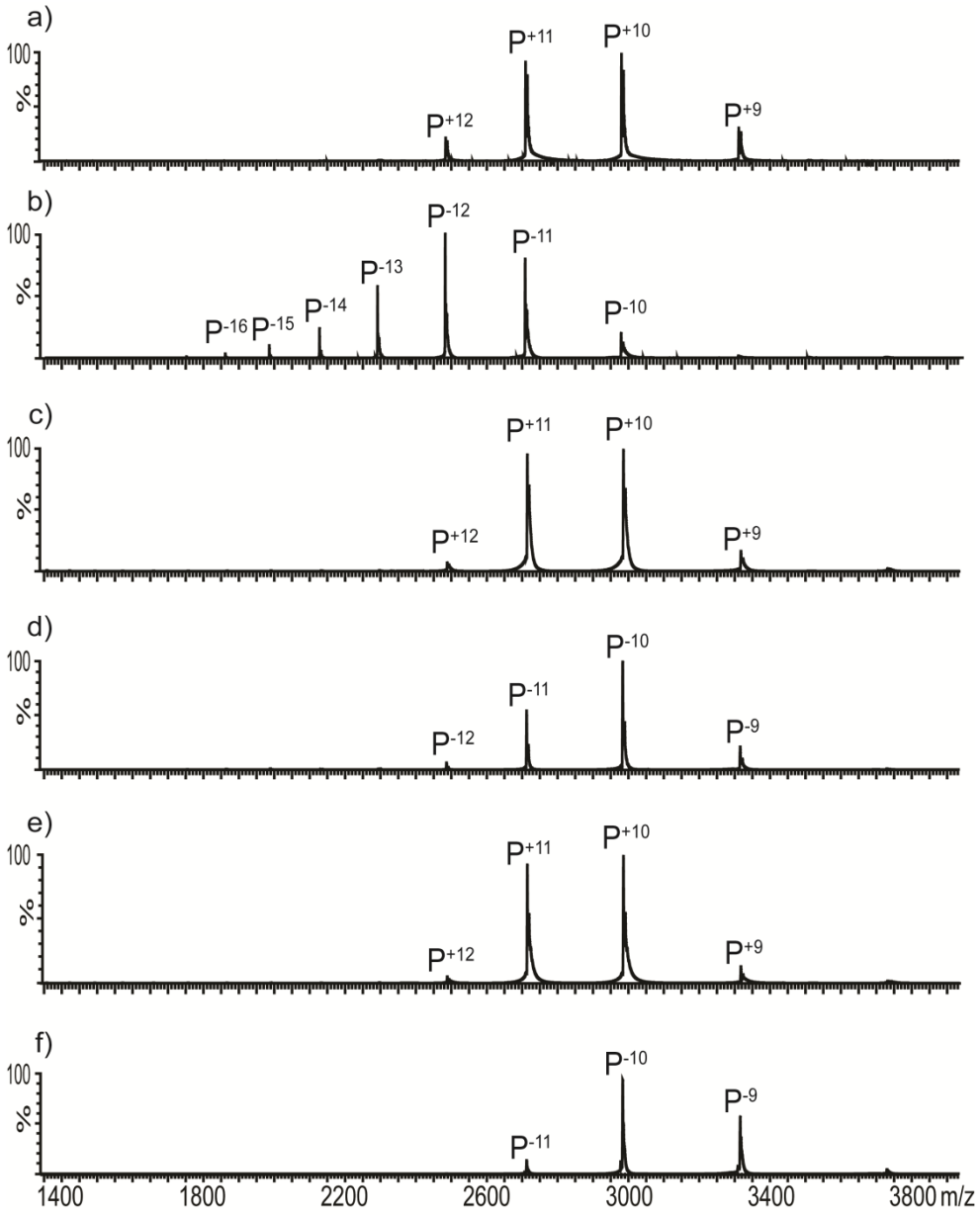
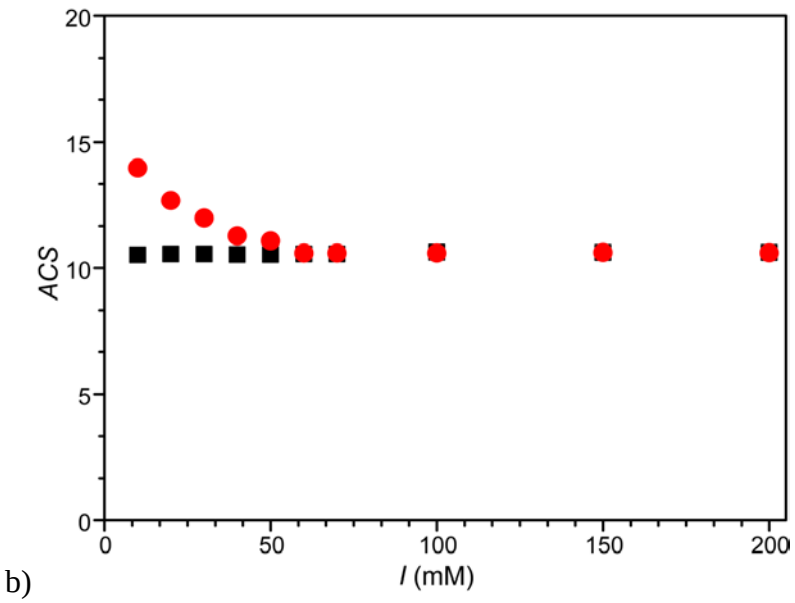


Figure 2.8 ESI mass spectra acquired for aqueous solutions of TcdB-B4B (P, 15 μM) with 10 mM ammonium acetate in (a) positive mode and (b) negative mode; 40 mM ammonium acetate in (c) positive mode and (d) negative mode; 200 mM ammonium acetate in (e) positive mode and (f) negative mode.

a)



b)

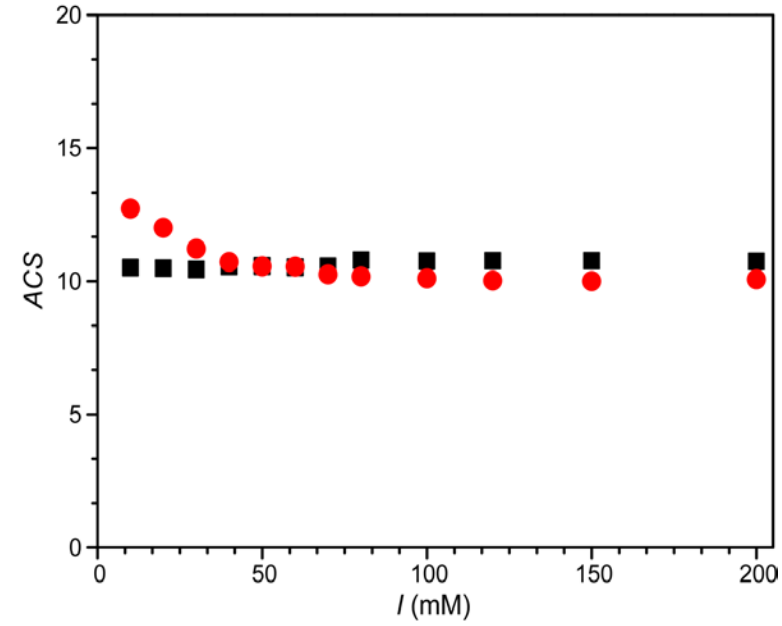


Figure 2.9 Plots of average charge state (*ACS*) versus ionic strength (*I*) measured from ESI mass spectra acquired in positive (■) and negative (●) ion mode for aqueous ammonium acetate solutions (pH 7) of two mutants (a) TcdB-B4A (10 μM) and (b) TcdB-B4B (10 μM). The ESI-MS measurements were carried out using identical instrumental/experimental conditions as those used for TcdB-B3C.

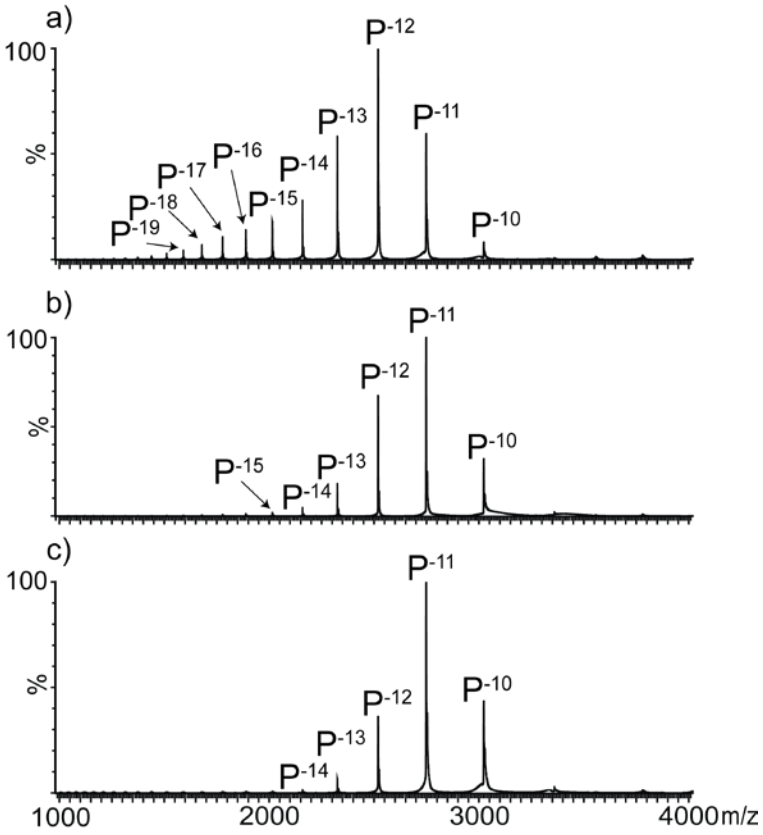


Figure 2.10 ESI mass spectra acquired for aqueous solutions of TcdB-B3C (P, 15 μM) with (a) 20 mM ammonium acetate; (b) 40 mM ammonium acetate and (c) 60 mM ammonium acetate.

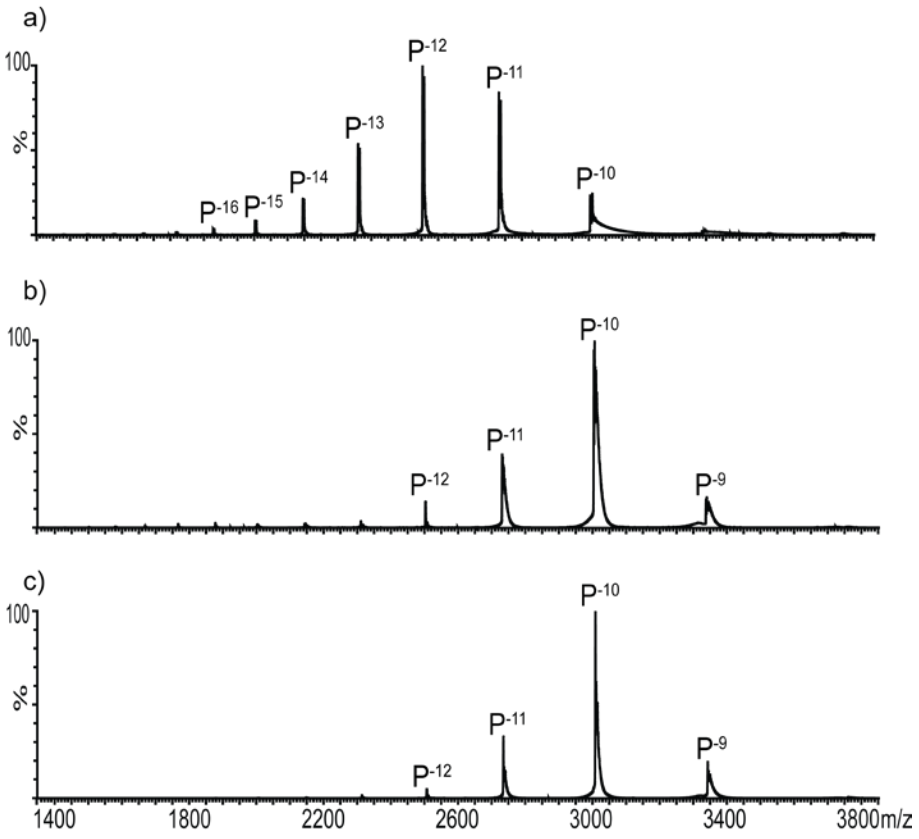


Figure 2.11 ESI mass spectra acquired for aqueous solutions of TcdB-B4A (P, 15 μM) with (a) 20 mM ammonium acetate; (b) 40 mM ammonium acetate and (c) 60 mM ammonium acetate in negative ion mode.

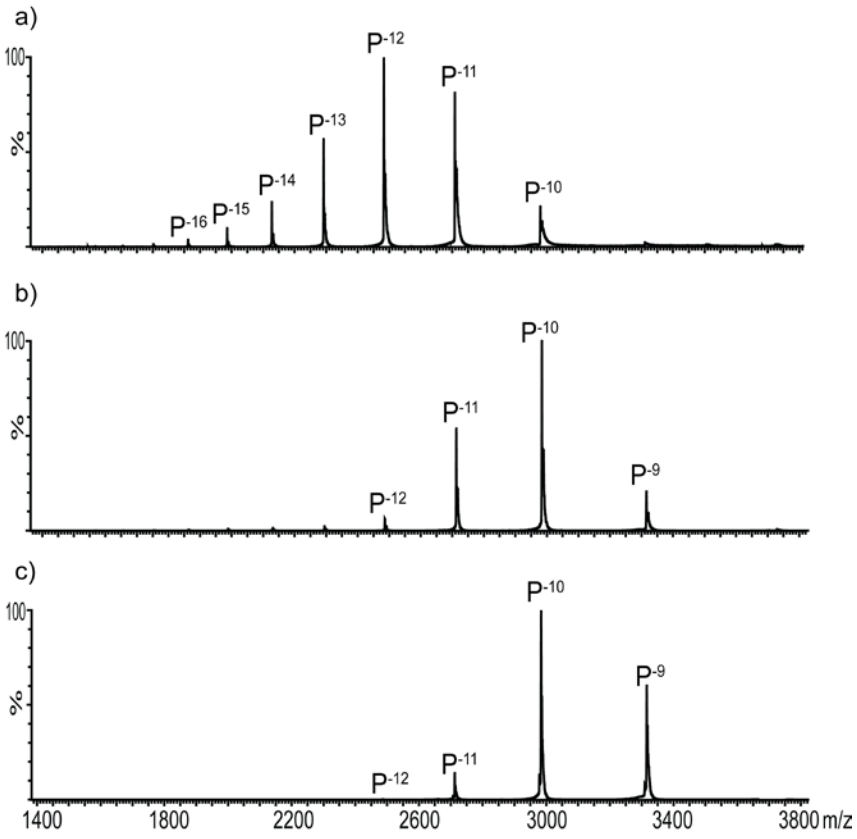


Figure 2.12 ESI mass spectra acquired for aqueous solutions of TcdB-B4B (P, 15 μM) with (a) 20 mM ammonium acetate; (b) 40 mM ammonium acetate and (c) 60 mM ammonium acetate in negative ion mode.

2.4 Conclusions

The results of the present study provide evidence for the occurrence of rapid, electrostatic-induced unfolding of acidic proteins in negatively charged ESI droplets. The extent of unfolding of the recombinant fragments B3C, B4A and B4B of TcdB in negative ion mode ESI-MS, which was monitored by changes in the CSD and the *ACS*, was found to be sensitive to the concentration of the solution "buffer", NH_4OAc. For neutral solutions of B3C at high *I*, >80 mM, the mass spectra exhibit a relatively narrow CSD and constant *ACS*, consistent with the protein having a compact structure. However, for solutions at lower *I*, the proteins exhibit a much broader CSD and a substantially larger (absolute) *ACS*, consistent with unfolding of the protein. In contrast, the CSD and *ACS* measured in positive ion mode are essentially independent of *I* (over the range investigated) and consistent with a folded protein. The results of ^{1}H NMR and CD spectroscopy and GFC measurements performed on solutions of B3C under low and high *I* conditions also suggest that the protein exists predominantly in a folded state in neutral aqueous solutions with *I* >10 mM. The results of ESI-MS measurements performed on a series of solutions of B3C with high *I* at pH 5 to 9 ruled out the possibility that the structural changes are related to ESI-induced changes in solution pH. Instead, it is proposed that the unfolding of B3C, observed in negative mode for solutions with low *I*, occurs in the ESI droplets and arises due to Coulombic repulsion between the negatively charged residues of the protein and droplet surface charge. The results of ESI-MS measurements performed on the mutants B4A and B4B, which contain fewer acidic residues than B3C, also

reveal a shift to higher absolute *ACS* at low *I*. However, in both cases the magnitude of the change is smaller than observed for B3C, consistent with the proposed electrostatic-induced unfolding mechanism.

2.5 Literature cited

1. Smith, R.D., J.A. Loo, C.G. Edmonds, C.J. Barinaga, and H.R. Udseth, *Anal. Chem.*, **1990**, 62, 882-899.

2. Loo, J.A., *Mass Spectrom. Rev.*, **1997**, 16, 1-23.

3. Loo, J.A., *Int. J. Mass spectrom.*, **2000**, 200, 175-186.

4. Benesch, J.L.P. and C.V. Robinson, *Curr. Opin. Struct. Biol.*, **2006**, 16, 245-251.

5. Heck, A.J.R., *Nat Meth*, **2008**, 5, 927-933.

6. Heck, A.J.R. and R.H.H. van den Heuvel, *Mass Spectrom. Rev.*, **2004**, 23, 368-389.

7. Konermann, L., J. Pan, and Y.-H. Liu, *Chem. Soc. Rev.*, **2011**, 40, 1224-1234.

8. Sharon, M. and C.V. Robinson, *Annu. Rev. Biochem*, **2007**, 76, 167-193.

9. Konermann, L. and D.J. Douglas, *Rapid Commun. Mass Spectrom.*, **1998**, 12, 435-442.

10. Kaltashov, I.A. and S.J. Eyles, *Mass Spectrom. Rev.*, **2002**, 21, 37-71.

11. Konermann, L., X. Tong, and Y. Pan, *J. Mass Spectrom.*, **2008**, 43, 1021-1036.

12. Powell, K.D., et al., *J. Am. Chem. Soc.*, **2002**, 124, 10256-10257.

13. Shoemaker, G.K., E.N. Kitova, and J.S. Klassen, *The Encyclopedia of Mass Spectrometry*, **2006**, 6, 810-822.

14. Konermann, L. and D.A. Simmons, *Mass Spectrom. Rev.*, **2003**, 22, 1-26.

15. Kitova, E., A. El-Hawiet, P. Schnier, and J. Klassen, *J. Am. Soc. Mass Spectrom.*, **2012**, 23, 431-441.

16. Liu, L., E. Kitova, and J. Klassen, *J. Am. Soc. Mass Spectrom.*, **2011**, 22, 310-318.

17. Marcsisin, S. and J. Engen, *Anal. Bioanal. Chem.*, **2010**, 397, 967-972.

18. Zhang, H., B.C. Gau, L.M. Jones, I. Vidavsky, and M.L. Gross, *Anal. Chem.*, **2011**, 83, 311-318.

19. Pan, Y., X. Ruan, M. Valvano, and L. Konermann, *J. Am. Soc. Mass Spectrom.*, **2012**, 23, 889-898.

20. Sohn, C.H., et al., *Anal. Chem.*, **2012**, 84, 2662-2669.

21. Sinz, A., *Mass Spectrom. Rev.*, **2006**, 25, 663-682.

22. Kaltashov, I.A. and A. Mohimen, *Anal. Chem.*, **2005**, 77, 5370-5379.

23. Testa, L., S. Brocca, and R. Grandori, *Anal. Chem.*, **2011**, 83, 6459-6463.

24. Chowdhury, S.K., V. Katta, and B.T. Chait, *J. Am. Chem. Soc.*, **1990**, 112, 9012-9013.

25. Grandori, R., *J. Mass Spectrom.*, **2003**, 38, 11-15.

26. Hamdy, O. and R. Julian, *J. Am. Soc. Mass Spectrom.*, **2012**, 23, 1-6.

27. Hall, Z. and C.V. Robinson, *J. Am. Soc. Mass Spectrom.*, **2012**, 1-8.

28. Konermann, L. and D. Douglas, *J. Am. Soc. Mass Spectrom.*, **1998**, 9, 1248-1254.

29. Kaltashov, I.A. and R.R. Abzalimov, *J. Am. Soc. Mass Spectrom.*, **2008**, 19, 1239-1246.

30. Ruotolo, B.T., et al., *Science*, **2005**, 310, 1658-1661.

31. Zhong, Y., S.J. Hyung, and B.T. Ruotolo, *Expert Rev Proteomics*, **2012**, 9, 47-58.

32. Duijn, E.v., A. Barendregt, S. Synowsky, C. Versluis, and A.J.R. Heck, *J. Am. Chem. Soc.*, **2009**, 131, 1452-1459.

33. Kitova, E.N., D.R. Bundle, and J.S. Klassen, *J. Am. Chem. Soc.*, **2002**, 124, 9340-9341.

34. Kitova, E.N., M. Seo, P.-N. Roy, and J.S. Klassen, *J. Am. Chem. Soc.*, **2008**, 130, 1214-1226.

35. Liu, L., D. Bagal, E.N. Kitova, P.D. Schnier, and J.S. Klassen, *J. Am. Chem. Soc.*, **2009**, 131, 15980-15981.

36. Scarff, C.A., K. Thalassinos, G.R. Hilton, and J.H. Scrivens, *Rapid Commun. Mass Spectrom.*, **2008**, 22, 3297-3304.

37. Jurneczko, E. and P.E. Barran, *Analyst*, **2011**, 136, 20-28.

38. Mirza, U.A. and B.T. Chait, *Int. J. Mass Spectrom. Ion Processes*, **1997**, 162, 173-181.

39. Breuker, K. and F.W. McLafferty, *Proc. Natl. Acad. Sci. U.S.A.*, **2008**, 105, 18145-18152.

40. Wyttenbach, T. and M.T. Bowers, *J. Phys. Chem. B*, **2011**, 115, 12266-12275.

41. Sterling, H., J. Prell, C. Cassou, and E. Williams, *J. Am. Soc. Mass Spectrom.*, **2011**, 22, 1178-1186.

42. Di Marco, V.B. and G.G. Bombi, *Mass Spectrom. Rev.*, **2006**, 25, 347-379.

43. Kharlamova, A., B.M. Prentice, T.-Y. Huang, and S.A. McLuckey, *Anal. Chem.*, **2010**, 82, 7422-7429.

44. Kharlamova, A., J. DeMuth, and S. McLuckey, *J. Am. Soc. Mass Spectrom.*, **2012**, 23, 88-101.

45. Kharlamova, A. and S.A. McLuckey, *Anal. Chem.*, **2011**, 83, 431-437.

46. Iavarone, A.T., J.C. Jurchen, and E.R. Williams, *Anal. Chem.*, **2001**, 73, 1455-1460.

47. Sterling, H., et al., *J. Am. Soc. Mass Spectrom.*, **2010**, 21, 1762-1774.

48. Sterling, H. and E. Williams, *J. Am. Soc. Mass Spectrom.*, **2009**, 20, 1933-1943.

49. Sterling, H.J., C.A. Cassou, A.C. Susa, and E.R. Williams, *Anal. Chem.*, **2012**, 84, 3795-3801.

50. Voth, D.E. and J.D. Ballard, *Clin. Microbiol. Rev.*, **2005**, 18, 247-263.

51. Curry, S.R., et al., *J. Clin. Microbiol.*, **2007**, 45, 215-221.

52. Ho, J.G.S., A. Greco, M. Rupnik, and K.K.-S. Ng, *Proc. Natl. Acad. Sci. U.S.A.*, **2005**, 102, 18373-18378.

53. Piotto, M., V. Saudek, and V. Sklenář, *J. Biomol. NMR*, **1992**, 2, 661-665.

54. Kuprowski, M.C. and L. Konermann, *Anal. Chem.*, **2007**, 79, 2499-2506.

55. Abzalimov, R.R., A.K. Frimpong, and I.A. Kaltashov, *Int. J. Mass spectrom.*, **2006**, 253, 207-216.

56. Van Berkel, G., K. Asano, and P. Schnier, *J. Am. Soc. Mass Spectrom.*, **2001**, 12, 853-862.

57. Wang, W., E.N. Kitova, and J.S. Klassen, *Anal. Chem.*, **2003**, 75, 4945-4955.

58. Wüthrich, K., NMR of Proteins and Nucleic Acids. 1986.

59. Page, R., W. Peti, I.A. Wilson, R.C. Stevens, and K. Wüthrich, *Proc. Natl. Acad. Sci. U.S.A.*, **2005**, 102, 1901-1905.

60. Hill, J.M., *Methods in molecular biology (Clifton, N.J.)*, **2008**, 426, 437-446.

61. Rehm, T., R. Huber, and T.A. Holak, *Structure*, **2002**, 10, 1613-1618.

62. Peti, W., et al., *J. Struct. Funct. Genomics*, **2004**, 5, 205-215.

63. Neri, D., M. Billeter, G. Wider, and K. Wuthrich, *Science*, **1992**, 257, 1559-1563.

64. Neri, D., G. Wider, and K. Wuthrich, *Proc. Natl. Acad. Sci. U.S.A.*, **1992**, 89, 4397-4401.

65. Tafer, H., S. Hiller, C. Hilty, C. Fernández, and K. Wüthrich, *Biochemistry*, **2004**, 43, 860-869.

66. Eliezer, D., *Methods in molecular biology (Clifton, N.J.)*, **2007**, 350, 49-67.

67. Mori, S., *Anal. Chem.*, **1989**, 61, 530-534.

68. Xu, D., J.C. Phillips, and K. Schulten, *J. Phys. Chem.*, **1996**, 100, 12108-12121.

69. Talaga, D.S. and J. Li, *J. Am. Chem. Soc.*, **2009**, 131, 9287-9297.

70. Wu, T., et al., *J. Appl. Phys.*, **2011**, 109.

Chapter 3

Quantifying Protein-Ligand Interactions by Direct ESI-MS Analysis. Evidence of Non-uniform Response Factors Induced by High Molecular Weight Molecules and Complexes*

3.1 Introduction

Noncovalent interactions between proteins and ligands (e.g. other proteins, carbohydrates, DNA or small molecules) are implicated in almost all biological processes, including cell signaling and recognition, inflammation, fertilization, infections by microbes and the immune response.[1-3] Investigations into the thermodynamics and kinetics of protein-ligand (PL) binding *in vitro* provide fundamental insights into biochemical reactions, serve to improve disease diagnosis and guide the development of new therapeutics. There are a number of established analytical methods available to quantify PL interactions, each with particular strengths and weaknesses. Among the most widely used methods for measuring the association constants (K_a) for PL complexes *in vitro* are isothermal titration calorimetry (ITC),[4-5] surface plasmon resonance (SPR) spectroscopy,[6] enzyme-linked immunosorbent assay,[7] and nuclear magnetic resonance spectroscopy.[8]

Recently, the direct electrospray ionization mass spectrometry (ESI-MS) assay has emerged as a powerful and versatile technique for detecting PL interactions in aqueous solution and quantifying their affinities.[9-13] The assay is

* A version of this chapter has been submitted for publication.

based on the direct detection of free and ligand-bound protein ions by ESI-MS analysis. The K_a for a given PL interaction is calculated from the ratio (*R*) of the total abundance (*Ab*) of ligand-bound and free protein ions measured for solutions with known initial concentrations of protein ($[P]_o$) and ligand ($[L]_o$). For example, the K_a for a 1:1 PL complex (eq 1) is calculated using eq 2:

$$P + L \rightleftharpoons PL \tag{1}$$

$$K_a = \frac{R}{[L]_o - \frac{R}{1+R}[P]_o} \tag{2}$$

where *R* is given by eq 3:

$$\frac{[PL]}{[P]} = \frac{Ab(PL)}{Ab(P)} = R \tag{3}$$

The assay has a number of analytical advantages over other binding assays, such as simplicity (no labeling or immobilization), speed (analysis takes typically 1-2 min), low sample consumption (<pmol per analysis), the ability to directly measure binding stoichiometry, analyze mixtures and measure multiple equilibria simultaneously.[14]

An underlying assumption of the direct ESI-MS assay is that the *R* ratio measured in the gas phase accurately reflects the concentration ratio of PL-to-P in solution (eq 3). The abundances of the gaseous P and PL ions measured by ESI-MS are related to their solution concentrations by response factors (*RF*), which collectively account for the ionization and detection efficiencies, eq 4:

$$\frac{[PL]}{[P]} = \frac{RF_P\, Ab(PL)}{RF_{PL}\, Ab(P)} = RF_{P/PL}\frac{Ab(PL)}{Ab(P)} \tag{4}$$

where RF_P and RF_{PL} are the response factors of P and PL, respectively, and $RF_{P/PL}$ is the ratio of the corresponding *RF* values (referred to as the relative response factor). Although the absolute *RF* values depend on many factors - the size and structure of P or PL, the solution composition and the instrumental parameters used for the measurements - uniform *RF*s for P and PL (i.e., $RF_{P/PL} \approx 1$) are expected in cases where L is small compared to P, such that the size and surface properties of the P and PL are similar.[15-16] Support for this assumption can be found in the similarity of K_a values measured by ESI-MS and other binding assays for a wide variety of PL complexes, including antibody-antigen, lectin-carbohydrate and enzyme-inhibitor complexes.[12,17-19]

Here, we report on the surprising finding that the presence of high molecular weight (MW) macromolecules or complexes in solution can influence $RF_{P/PL}$ values and, as a result, introduce significant errors to ESI-MS affinity measurements. The unexpected effect of large solute on $RF_{P/PL}$ values came to light in a recent ESI-MS study of carbohydrate interactions with viral protein particles. These measurements were carried out using the proxy protein method, in which ligand binding to the target protein is quantified using the direct ESI-MS assay to monitor L binding to a proxy protein (P_{proxy}).[20]

3.2 Experimental section

3.2.1 Proteins and ligands

The carbohydrate-binding single chain variable fragment (scFv, MW 26 539 Da) of the monoclonal antibody Se155-4 was produced and purified as described elswhere.[21] The scFv was concentrated and dialyzed against aqueous 50 mM ammonium acetate using MICROSEP microconcentrators (Millipore Corp., Bedford, MA, USA) with a molecular weight cutoff of 10 kDa and stored at 4 ^{o}C, if not used immediately. A truncated version of wildtype P22 bacteriophage homotrimeric tailspike protein (TSP, MW 180 kDa) was a gift from Prof. C. Szymanki (University of Alberta). The P particle of norovirus VA387 was a gift from Prof. X. Jiang (University of Cincinnati College of Medicine). The octasaccharide (**O**, [α-D-Gal*p*-(1→2)-[α–D-Abe*p*-(1→3)]-α-D-Man*p*-(1→4)-α-L-Rha*p*]$_2$) and trisaccharide (**T**, Methylα-D-Tal*p*-(1→2)-[α-D-Abe*p*-(1→3)]-α-Man*p*] ligands were gifts from Prof. D. Bundle (University of Alberta). Dextran polysaccharides, with average MWs of 100 kDa, 500 kDa and 2000 kDa, were purchased from Sigma-Aldrich Canada (Oakville, Canada). Surfactant *n*-dodecyl-β-D-maltoside (DDM) was purchased from Thermo Fisher Scientific Inc. (Toronto, Canada) and the bovine monosialotetrahexosylganglioside (β-D-Gal*p*-(1→3)-β-D-Gal*p*NAc-(1→4)[α-D-Neu5Ac-(2→3)]-β-D-Gal*p*-(1→4)-β-D-Glc*p*-ceramide, GM1) was purchased from Axxora LLC (San Diego, CA). Recombinant membrane protein MSP1E1 (MW 27 494 Da) was prepared using plasmid pMSP1E1 acquired from Addgene (Cambridge, MA). Protein expression and purification was then carried out using the procedure described at

http://sligarlab.life.uiuc.edu/nanodisc.html. 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC, MW 677.9 Da) dissolved in chloroform was purchased from Avanti Polar Lipids (Alabaster, AL). Nanodiscs (ND) composed of DMPC were prepared using procedures described previously.[22-23] Stock solutions of each oligosaccharide and Dextran polysaccharides were prepared by dissolving a known amount of the solid sample in ultrafiltered water (Milli-Q, Millipore) to yield a final concentration of 1 mM. The solutions were stored at -20 °C until needed. A ND stock solution was concentrated and dialyzed against 200 mM ammonium acetate (pH 7.0) using an amicon microconcentrator with a MW cut-off of 30 kDa and stored at -80 °C until needed. The concentration of the ND solutions was determined by absorbance at 280 nm using the extinction coefficient of MSP1E1. Stock solutions of each surfactant was prepared by dissolving a known amount of the sample in ultrafiltered water (Milli-Q, Millipore) to yield a final concentration which is smaller than its CMC value. The solutions were stored at -20 °C until needed.

3.2.2 Mass spectrometry

The binding measurements were carried out at room temperature using a 9.4T ApexQe Fourier transform ion cyclotron resonance (FTICR) mass spectrometer (Bruker, Billerica, MA) and a Synapt G2 quadrupole-ion mobility separation-time of flight (Q-IMS-TOF) mass spectrometer (Waters, UK). In both cases, nanoflow ESI (nanoESI) source was equipped. To perform nanoESI, tips were produced from borosilicate tubes (1.0 mm o.d., 0.68 mm i.d.), pulled to ~5 μm o.d. at one

end using a P-97 micropipette puller (Sutter Instruments, Novato, CA). A platinum wire was inserted into the nanoESI tip, and a capillary voltage was applied to carry out ESI.

ApexQe 9.4T FTICR mass spectrometer. The droplets and gaseous ions produced by ESI were introduced into the mass spectrometer through a metal sampling capillary (0.5 mm i.d.). Nitrogen gas at a flow rate of 2.0 L min^{-1} and 90 ^{o}C was used as a drying gas. The capillary entrance and exit voltages were both held at 280 V. A deflector voltage of 225 V was used. Gaseous ions were transmitted through the first funnel and skimmer held at 150 V and 20 V, respectively, and then through the second funnel and skimmer held at 7.6 V and 5.3 V, respectively. The ions were stored electrodynamically in an rf hexapole for 0.5 s and then further accumulated in a hexapole collision cell for 0.4 s. Following accumulation, the ions were transferred into the ion cell. The front and back trapping plates of the cell were maintained at 0.9 and 1.0 V, respectively, throughout the experiment. The typical base pressure for the instrument was ~1 X 10^{10} mbar. Data acquisition and analysis were performed using ApexControl, version 4.0 (Bruker Daltonics). A minimum of 30 transients with 32k data points per transient were used for each acquisition.

Synapt G2 Q-IMS-TOF mass spectrometer. Mass spectra were obtained in positive ion modes using cesium iodide (concentration 30 ng μL^{-1}) for calibration. Given below are some of the instrumental conditions used to carry out the measurements in positive ion mode. A capillary voltage of 1.0-1.3 kV was applied to carry out nanoESI. A cone voltage of 35 V was used and the source block

temperature was maintained at 70 ºC. Other important voltages for ion transmission, that is the injection voltages into the trap and transfer ion guides, were maintained at 5 V and 2 V, respectively. Argon was used in the trap and transfer ion guides at a pressure of 2.22 x 10^{-2} mbar and 3.36 x 10^{-2} mbar, respectively. Data acquisition and processing were carried out using MassLynx (v 4.1).

3.2.3 Solvent Accessible Surface (SAS) Area Calculations

The solvent accessible surface (SAS) areas were calculated for the free proteins (scFv and TSP) and free ligands (**O** and **T**) separately using msroll and the Lee and Richards algorithm.[24] The structure of scFv was generated based on the reported crystal structure for scFv complexed with its native trisaccharide ligand (α-D-Gal*p*-(1→2)-[α-D-Abe*p*-(1→3)]-α-Man*p*), 1MFA,[21] using Pymol.[25] The structures of the TSP trimer and **O** were obtained from the reported crystal structure of TSP complexed with **O**, 1TYX, using Pymol.[26] The structure of **T** (methylα-D-Tal*p*-(1→2)-[α-D-Abe*p*-(1→3)]-α-Man*p*) was generated using the GLYCAM oligosaccharide on-line builder.[27] The probe radius was set to 1.40 Å. For the purposes of estimating the change in the protein SAS, the ligand was treated as a planar molecule such that the reduction in protein SAS is equal to one half of the ligand SAS value.

3.3 Results and Discussion

It was found that, for some $P_{proxy}L$ interactions, *R* decreased in a manner that was inconsistent with competitive protein binding.[20] As an example, shown in Figure 3.1 are ESI mass spectra measured for solutions containing a single chain variable fragment (scFv, MW 26 539 Da) of the monoclonal antibody Se155-4 and the octasaccharide ligand (**O**, [α-D-Gal*p*-(1→2)-[α-D-Abe*p*-(1→3)]-α-D-Man*p*-(1→4)-α-L-Rha*p*]$_2$) in the absence and presence of a P particle derived from the norovirus VA387 strain.[28] The spherical P particle has a MW of 865 253 Da and a diameter of ~20 nm.[29] It was found that the addition of the P particle to the solution resulted in a decrease in the abundance of the (scFv + **O**) complex relative to scFv (Figures 3.1a and 3.1b). The dependence of *R* on the concentration of P particle is shown in Figure 3.2. It can be seen that *R* initially decreased with increasing P particle concentration but reached a constant value at concentrations ≥ 5 μM. In the absence of P particle, the measured *R* value corresponds to an affinity of $(5.4 \pm 0.1) \times 10^5$ M^{-1}, which is in good agreement with a value determined by isothermal titration calorimetry ($(5.3 \pm 1.6) \times 10^5$ M^{-1}).[30] The decrease in *R* caused by addition of 6 μM P particle corresponds to a 63% decrease in the apparent K_a for the (scFv + **O**) complex.

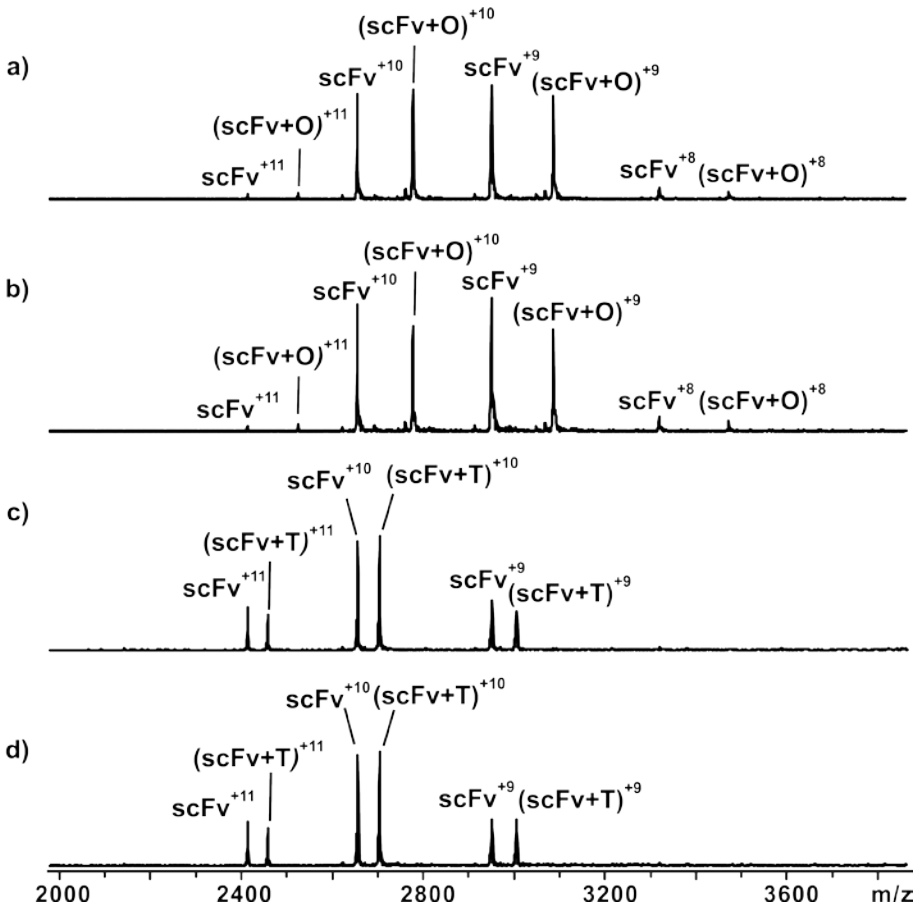


Figure 3.1 ESI mass spectra obtained in positive ion mode for aqueous ammonium acetate (10 mM, pH 6.9) solutions of scFv (6.5 μM) and (a) **O** (5.0 μM), (b) **O** (5.0 μM) and P particle (2.0 μM), (c) **T** (8.0 μM), and (d) **T** (8.0 μM) and P particle (2.0 μM).

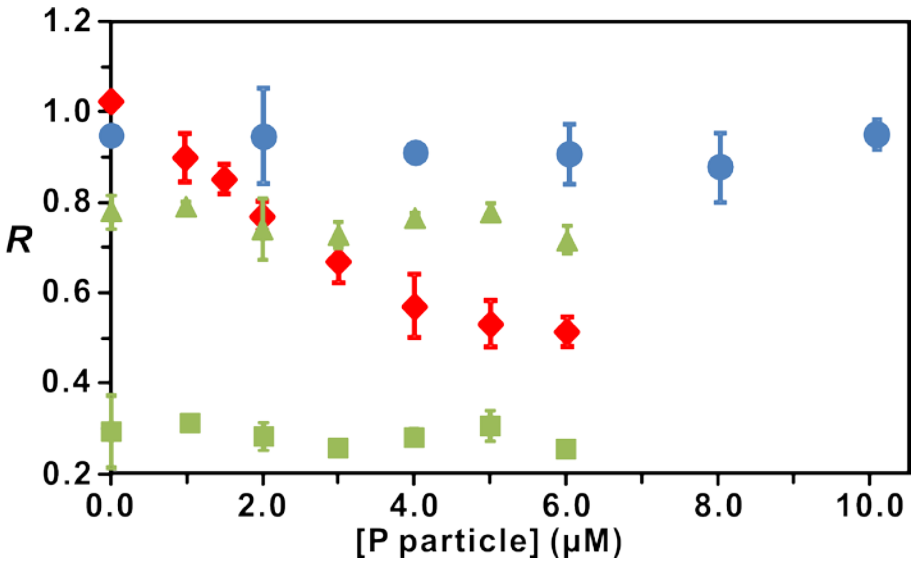


Figure 3.2 Plots of *R* versus P particle concentration measured for aqueous ammonium acetate (10 mM, pH 6.9) solutions of scFv (6.5 μM) with **O** (5.0 μM) or **T** (8.0 μM) and TSP (1.7 μM) with **O** (2.0 μM). The *R* values for (scFv+**O**)/scFv, (scFv+**T**)/scFv, (TSP+**O**)/TSP and (TSP+2**O**)/TSP are represented by ♦, ● and ▲ and ■, respectively.

The initial decrease in *R* with increasing P particle concentration is qualitatively consistent with competitive binding of **O** to scFv and the P particle. However, the overall concentration dependence of *R* can't be explained quantitatively on the basis of competitive protein binding. Indeed, analogous measurements performed using the 180 kDa homotrimeric tailspike protein (TSP)

of P22 bacteriophage, which can bind up to three molecules of **O**,[20,30] in place of the scFv revealed that the addition of P particle had no significant effect on the measured *R* values, Figures 3.2 and 3.3. This result excludes the possibility that **O** binds to the P particle in solution and, instead, suggests that the influence of the P particle on the ratio of bound (to **O**) and unbound scFv is due to differential *RF*s arising from differences in the surface activity of P and PL.

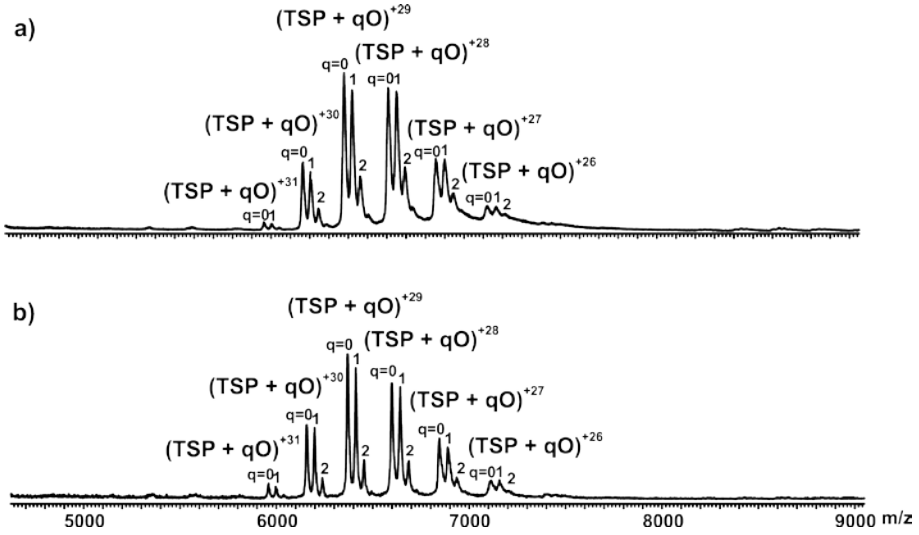


Figure 3.3 ESI mass spectra measured for an aqueous ammonium acetate (10 mM, pH 6.9) solution of (a) TSP (1.7 μM) and **O** (2.0 μM) and (b) TSP (1.7 μM), **O** (2.0 μM) and P particle (2.0 μM).

Based on solvent accessible surface (SAS) area calculations, it is estimated that binding of **O** to scFv results in the replacement of 5.8% of the protein surface with that of the more hydrophilic oligosaccharide, results shown in Table 3.1. This contrasts with a change of only 1.4% for the TSP bound to one **O** (and 2.9%

for binding to two **O**). Further support for the hypothesis that the reduction in *R* measured for the scFv/(scFv + **O**) system upon introduction of P particle to solution is due to differential changes in the surface activities of the bound and unbound forms of the protein comes from the complete absence of an effect of the P particle on the apparent binding equilibrium between scFv and the trisaccharide ligand (**T**, methylα-D-Tal*p*-(1$\rightarrow$2)-[α-D-Abe*p*-(1$\rightarrow$3)]-α-D-Man*p*) (Figures 3.1c and 3.1d). In this case, binding of **T** to scFv is estimated to result in a change of only 2.9% in the protein SAS.

Table 3.1 Solvent accessible surface area (SAS) calculated for scFv and TSP and the carbohydrate ligands **O** and **T** and the estimated change in protein SAS (ΔSAS) upon ligand binding.

Protein/Ligand	SAS ($Å^2$)	T binding ΔSAS (%)[a]	O binding ΔSAS (%)[a]
scFv	11691	2.9	5.8
TSP	47254	NA [b]	1.4 (2.9) [c]
T	685	-	-
O	1367	-	-

a. Estimated assuming carbohydrate ligands are planar molecules such that half of the ligand surface area is in contact with the protein in the bound form. b. Not applicable. c. Change in SAS corresponding to binding of two molecules of **O**.

Further investigation into the influence of solution composition on ESI-MS binding measurements revealed that the addition of macromolecules and large

non-covalent assemblies to solution generally influences the relative abundances of PL and P, with the effect being more pronounced as the differences between the surface properties of the P and PL complexes become more significant. For example, the concentration effects of the branched glucan polysaccharide dextran, with average MWs of 100, 500 and 2000 kDa and Stokes radii of 6.9, 14.7 and 27 nm,[31] respectively, on the *R* values measured for the (scFv + **T**) and (scFv + **O**) complexes are shown in Figure 3.4. Illustrative ESI mass spectra are shown in Figures 3.5-3.8 for solutions containing the 100 kDa and 2000 kDa dextran polysaccharide. In all cases, the *R* values measured for the (scFv + **T**) complex were found to be essentially independent of polysaccharide concentration (Figure 3.4). Furthermore, the 100 kDa polysaccharide had no effect on the *R* values measured for the (scFv + **O**) complex (Figure 3.4a). In contrast, introduction of the 500 kDa or 2000 kDa polysaccharide to solution resulted in a decrease in *R* (Figure 3.4b and 3.4c). The magnitude of the effect was sensitive to the MW of the polysaccharide, with the 2000 kDa polysaccharide inducing a more significant decrease in *R* at a given concentration. Similar to the behavior observed upon addition of the P particle to solution, *R* measured in the presence of the 2000 kDa polysaccharide was essentially constant at concentrations >4 μM.

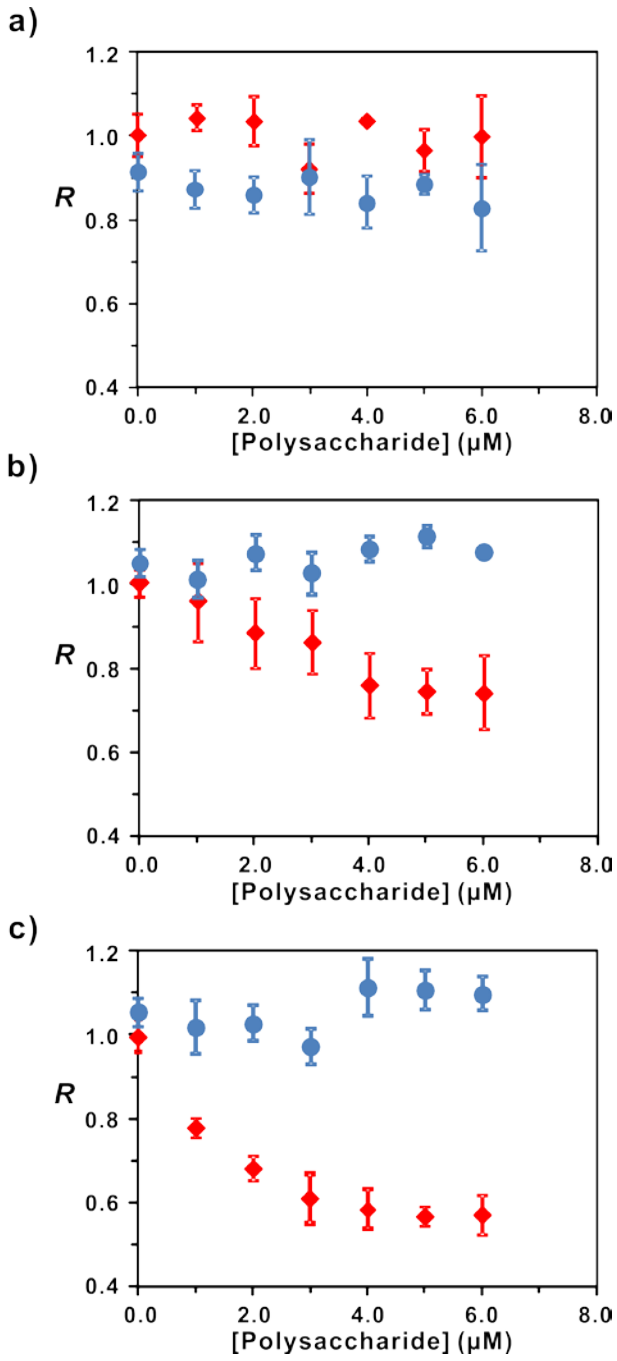


Figure 3.4 Plots of *R* versus concentration of dextran polysaccharide with average MW of (a) 100 kDa (b) 500 kDa and (c) 2000 kDa, measured for aqueous ammonium acetate (10 mM, pH 6.9) solutions of scFv (6.5 μM) and **O** (5.0 μM) or **T** (8.0 μM). The *R* values for (scFv+**O**)/scFv and (scFv+**T**)/scFv are represented by ♦ and ●, respectively.

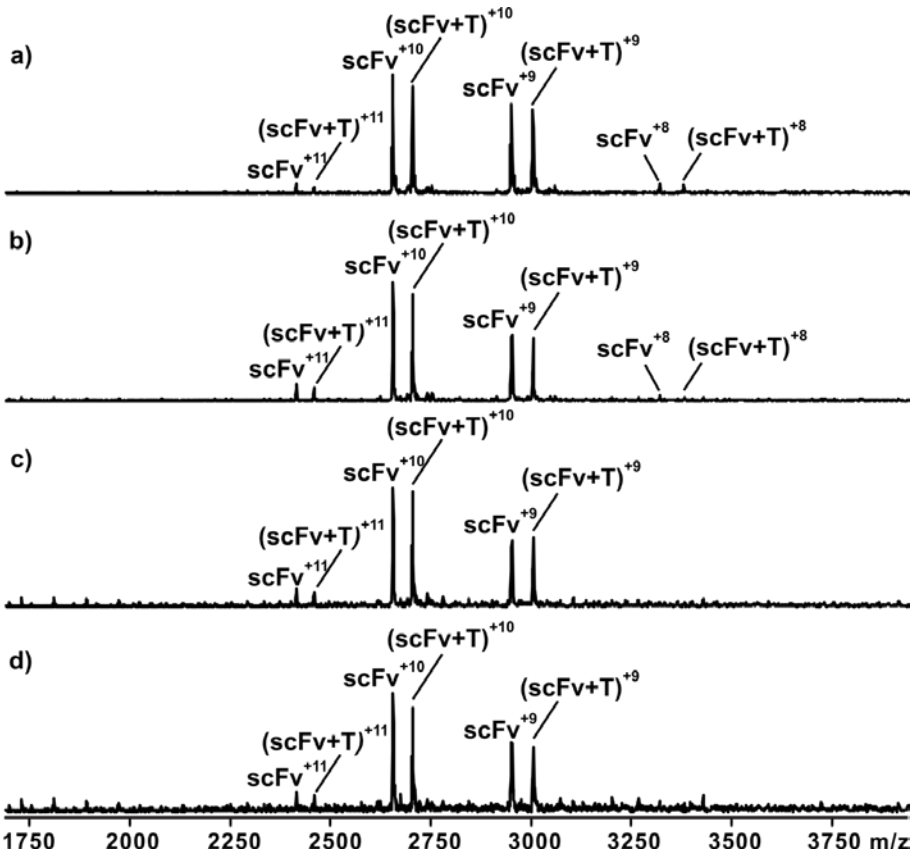


Figure 3.5 ESI mass spectra measured for an aqueous ammonium acetate (10 mM, pH 6.9) solution of scFv (6.5 μM), **T** (8.0 μM) and 100 kDa dextran polysaccharide at a concentration of (a) 0.0 μM, (b) 2.0 μM, (c) 4.0 μM and (d) 6 μM.

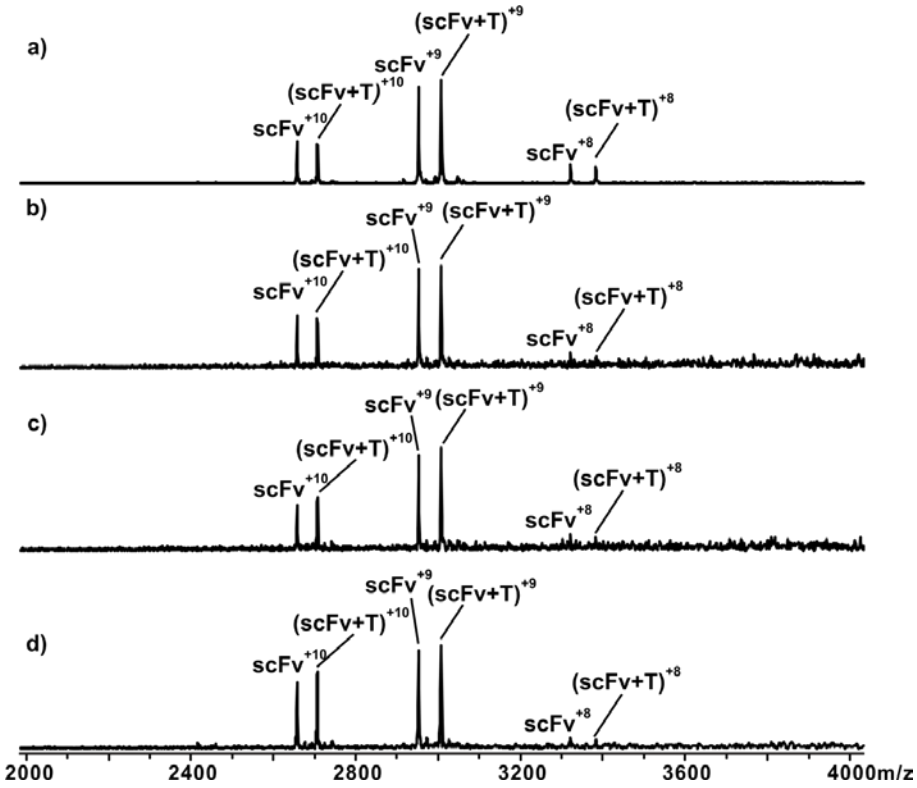


Figure 3.6 ESI mass spectra measured for an aqueous ammonium acetate (10 mM, pH 6.9) solution of scFv (6.5 μM), **T** (8.0 μM) and 2000 kDa dextran polysaccharide at a concentration of (a) 0.0 μM, (b) 2.0 μM, (c) 4.0 μM and (d) 6 μM.

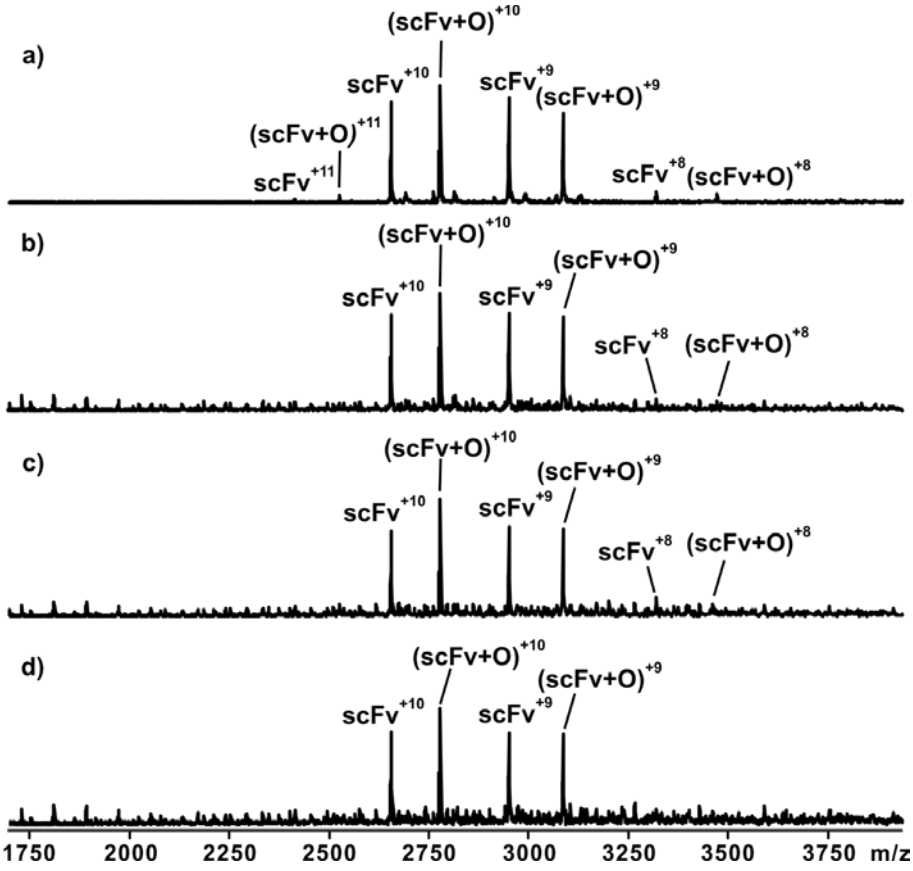


Figure 3.7 ESI mass spectra measured for an aqueous ammonium acetate (10 mM, pH 6.9) solution of scFv (6.5 μM), **O** (5.0 μM) and 100 kDa dextran polysaccharide at a concentration of (a) 0.0 μM, (b) 2.0 μM, (c) 4.0 μM and (d) 6 μM.

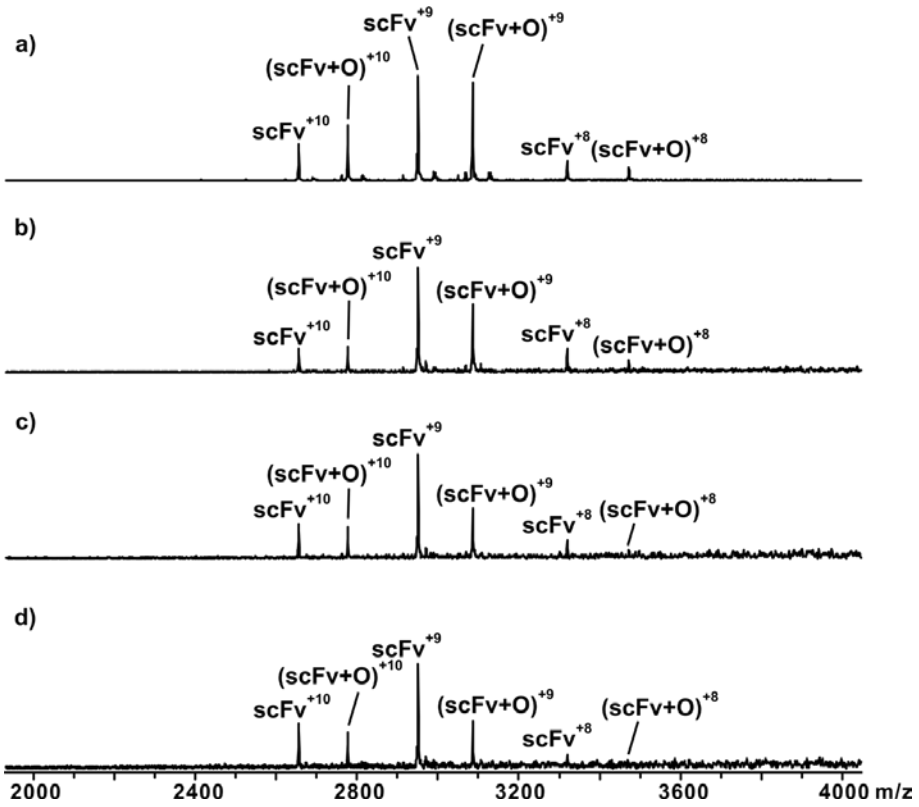


Figure 3.8 ESI mass spectra measured for an aqueous ammonium acetate (10 mM, pH 6.9) solution of scFv (6.5 µM), **O** (5.0 µM) and 2000 kDa dextran polysaccharide at a concentration of (a) 0.0 µM, (b) 2.0 µM, (c) 4.0 µM and (d) 6 µM.

The influence of a nanodisc (ND),[32] which is a discoidal phospholipid bilayer surrounded by two copies of an amphipathic scaffold protein that solubilize the lipid bilayer, on the relative *RF*s for the (scFv + **O**) and (scFv + **T**) interactions was also investigated. The ND used in the present study, which was

composed of the phospholipid DMPC and recombinant membrane scaffold protein MSP1E1, has an estimated MW of 190 kDa and Stokes radius of 10.4 nm.[32-33] As can be seen from Figure 3.9, the addition of ND to solution had no effect on the *R* values measured for the (scFv + **T**) complex. In contrast, the *R* values measured for the (scFv + **O**) complex initially decreased with increasing ND concentration, and reached a constant value at concentrations >3 μM. Notably, the change in *R* (~60%) at the higher concentrations is comparable to that observed for the P particle and the 2000 kDa dextran.

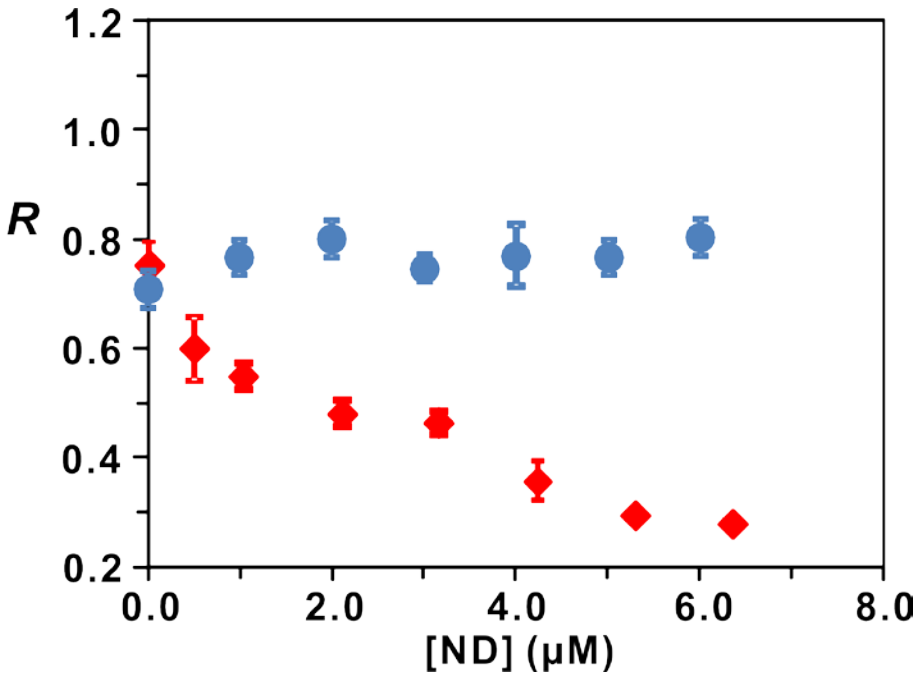


Figure 3.9 Plots of *R* versus nanodisc (ND) concentration measured for aqueous ammonium acetate (10 mM, pH 6.9) solutions of scFv (6.5 μM) and **O** (5.0 μM) or **T** (8.0 μM). The *R* values for (scFv+**O**)/scFv and (scFv+**T**)/scFv are represented by ♦ and ●, respectively.

It is important to note that the influence of large and high MW solute on the relative *RF* values for the scFv/(scFv + **O**) and scFv/(scFv + **T)** systems is independent of the MS instrumentation used to acquire the ESI mass spectra. For example, shown in Figure 3.10 and 3.11 are plots of *R* versus concentration of the 2000 kDa dextran and ND, respectively, acquired using a Synapt G2 quadrupole-ion mobility separation-time of flight mass spectrometer (Waters, UK). These can be compared to the data shown in Figure 3.4c and 3.9, respectively, which were acquired using a 9.4 T ApexQe FTICR mass spectrometer (Bruker, Billerica, MA). It can be seen that the concentration dependence of *R*, measured using the two different instruments, is qualitatively similar in the two cases.

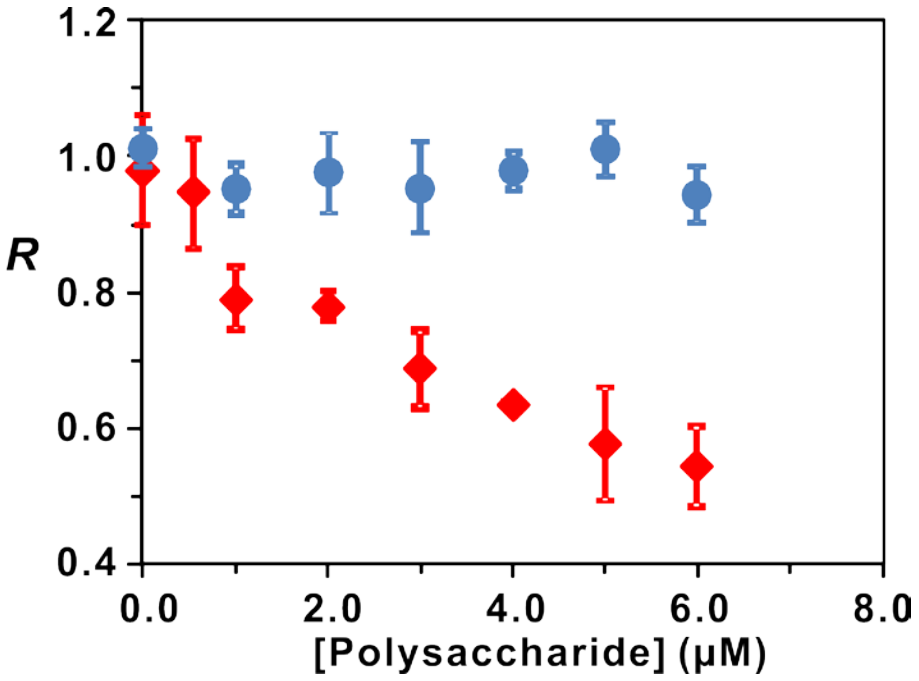


Figure 3.10 Plots of *R* versus concentration of 2000 kDa dextran polysaccharide measured for aqueous ammonium acetate (10 mM, pH 6.9) solutions of scFv (6.5 μM) and **O** (5.0 μM) or **T** (8.0 μM) using a Waters Synapt G2 mass spectrometer. The *R* values for (scFv+**O**)/scFv and (scFv+**T**)/scFv are represented by ♦, and ● respectively.

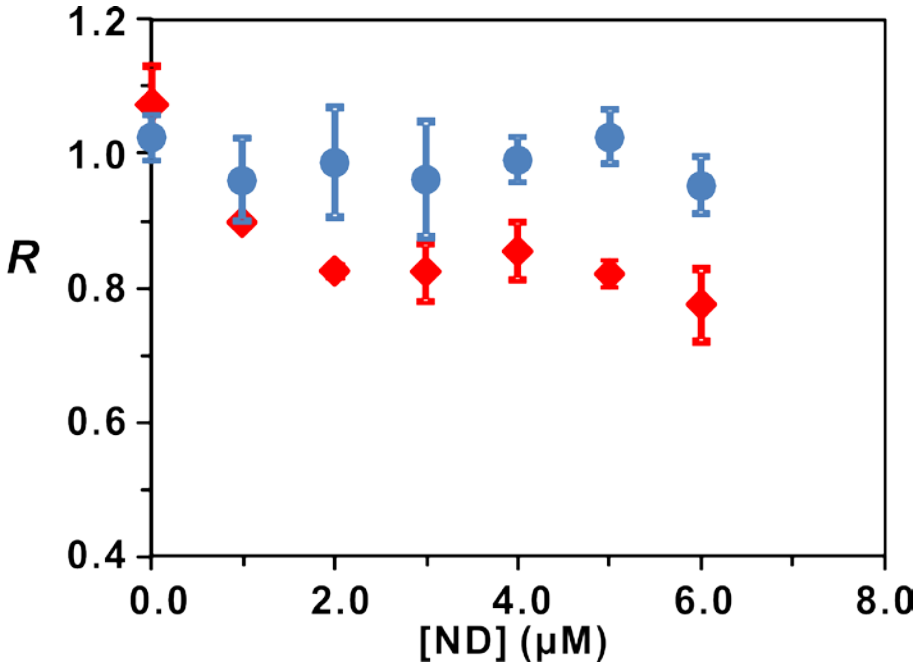


Figure 3.11 Plots of *R* versus concentration of nanodisc (ND) measured for aqueous ammonium acetate (10 mM, pH 6.9) solutions of scFv (6.5 μM) and **O** (5.0 μM) or **T** (8.0 μM) using a Waters Synapt G2 mass spectrometer. The *R* values for (scFv+**O**)/scFv and (scFv+**T**)/scFv are represented by ♦, and ● respectively.

The aforementioned results clearly demonstrate that high MW polymers and complexes can alter the *R* values for protein-carbohydrate interactions and, thereby, influence the affinity measured by ESI-MS. But what is the physical origin of this effect? To answer this question it is useful to first summarize the key findings of this study: (1) The influence of solute on *R* is more pronounced in

cases where the surface properties of the P and PL complexes are significantly different. (2) The magnitude of the change in *R*, at a given solute concentration, appears to correlate with the size of the solute and the difference in SAS of the free and ligand-bound protein. (3) The change in *R* appears to reach a maximum value at solute concentrations of between 3 - 5 μM. Taken together, the first two findings suggest that the changes in *R* reflect the presence of one or more solute molecules in the ESI droplets, which reduces the number of surface sites available to P and PL. When P is more hydrophobic than PL (which is the case for the protein-carbohydrate complexes investigated in the present study), it will out-compete PL for the available surface sites, thereby reducing the measured *R* value. The magnitude of the effect is expected, therefore, to be related to the fraction of droplets that contain high MW solute. The distribution of solute in the ESI droplets (as a function of solute concentration) is expected to be Poisson in nature. Assuming an initial droplet diameter of 100 nm,[34-35] the distributions expected at 1, 5 and 10 μM solute concentrations are shown in Figure 3.12. It can be seen that, at a concentration of 1 μM, the majority (73%) of the initial ESI droplets will contain no solute molecules at all. However, that value rapidly increases with increasing concentration and, at 5 μM, more than 79% of the droplets will contain one or more solute. The results of this analysis provide a qualitative explanation for the initial concentration dependence observed for the *R* values. However, they do not, on their own, explain why the *R* values appear to reach a limiting value at solute concentrations of 3 – 5 μM. It is possible that this effect is related to physical limitations in the number of solute molecules that can be accommodated

in the ESI droplets. The ESI droplets that ultimately produce the gas-phase protein ions may be too small to accommodate more than a single, large solute molecule. In this case, increasing the solute concentration beyond a certain value would have no influence on the measured *R* values.

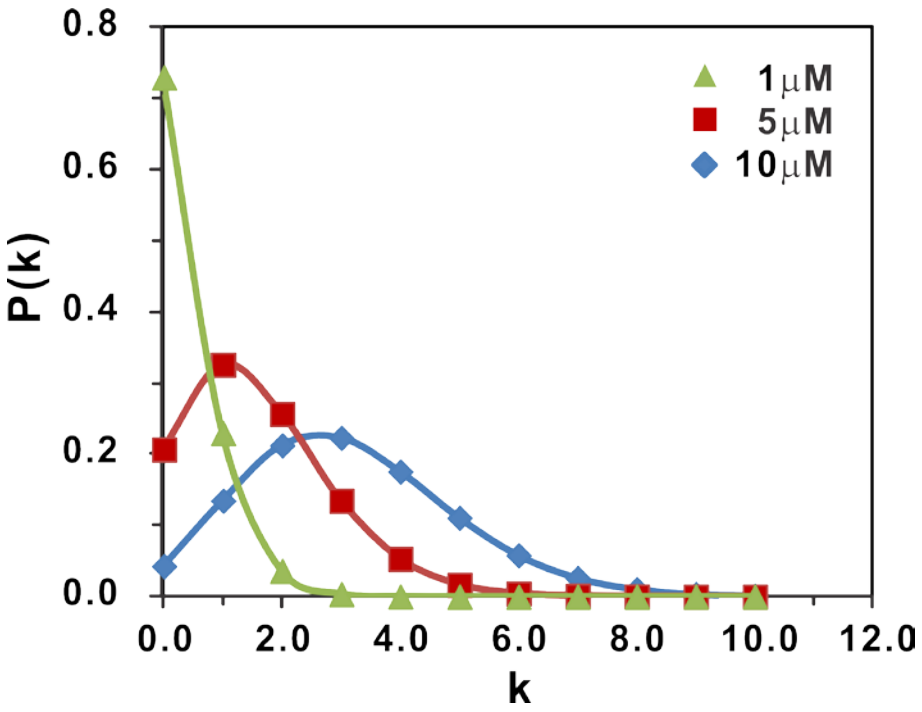


Figure 3.12 Poisson distribution of solute in the ESI droplets assuming an initial droplet diameter of 100 nm and a solute concentration of 1 μM (▲), 5 μM (■) and 10 μM (♦). k is the number of molecules per droplet and P(k) is the probability the droplet contains k solute molecules.

As an additional test of the hypothesis that the non-uniform *RF* values are due to the loss of surface sites in the ESI droplets and enhanced competition between P and PL for the available sites, ESI measurements were performed on

solutions of scFv and **O** or **T**, in the presence of surfactants (*n*-dodecyl-β-D-maltoside (DDM) and the ganglioside GM1) at different concentrations. As can be seen from Figure 3.13, the addition of DDM had no effect on the *R* values measured for (scFv + **T**), while the *R* values measured for (scFv + **O**) exhibit a continuous decrease with increasing DDM concentration (up to concentrations >15μM). These findings are consistent with the hypothesis that surface active compounds effectively block surface sites in the ESI droplets and support the proposed explanation that the effect of large solute on the *R* values measured for the (scFv + **O**) complex is due to competition between the ligand-bound and unbound forms of scFv for the available surface sites. Moreover, the distinct DDM concentration dependence of *R* measured for (scFv + **O**) can be taken as qualitative support for the suggestion that the observations of constant *R* values at higher concentrations of large solute for (scFv + **O**) are due the inability of the ESI droplets to accommodate more than a single, large solute molecule.

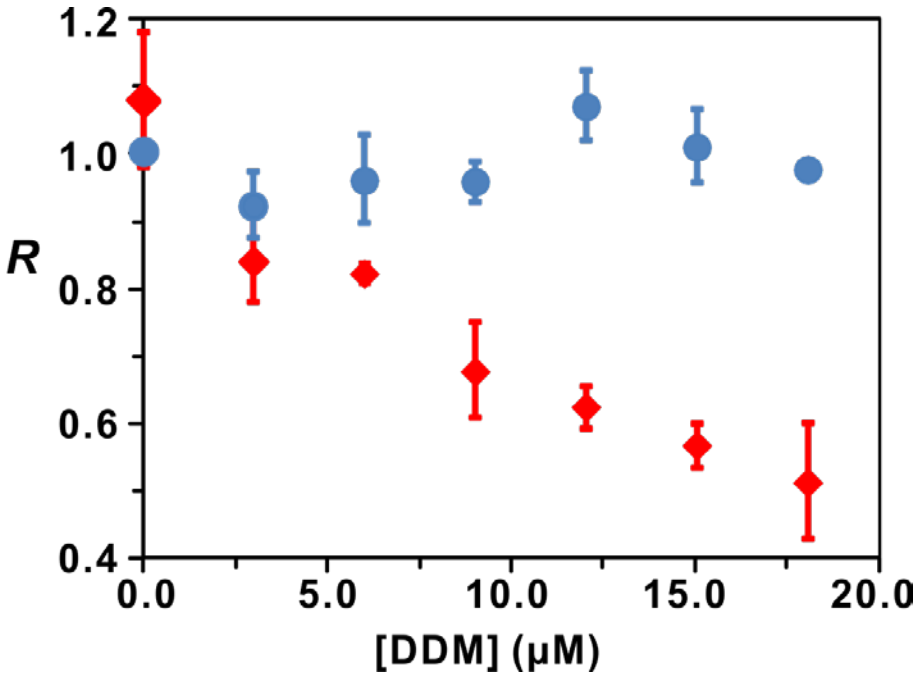


Figure 3.13 Plots of *R* versus concentration of DDM measured for aqueous ammonium acetate (10 mM, pH 6.9) solutions of scFv (6.5 μM) and **O** (5.0 μM) or **T** (8.0 μM). The *R* values for (scFv+**O**)/scFv and (scFv+**T**)/scFv are represented by ♦, and ● respectively.

In contrast to the behavior observed for DDM, the addition of GM1 to solution produced results that are very similar to those found for the large solute investigated (Figure 3.14). Gangliosides, such as GM1, are known to readily form micelles in solution.[29] For example, the critical micelle concentration of GM1 (~20 μM), is approximately one tenth that of DDM (~0.2 mM).[29,30] Therefore, it is likely that GM1 rapidly self-assembles into micelles or micelle-like complexes

during the ESI process due to solvent evaporation (and concomitant concentration increase), and thereby effectively behaves similar to large solute, as opposed to monomeric surfactant.

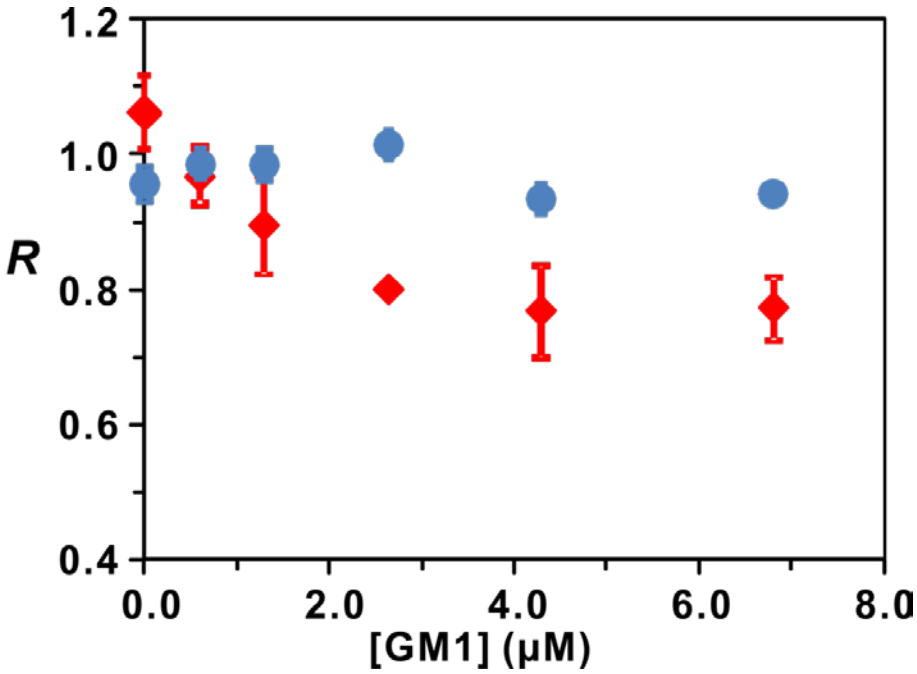


Figure 3.14 Plots of *R* versus concentration of GM1 measured for aqueous ammonium acetate (10 mM, pH 6.9) solutions of scFv (6.5 μM) and **O** (5.0 μM) or **T** (8.0 μM). The *R* values for (scFv+**O**)/scFv and (scFv+**T**)/scFv are represented by ♦, and ● respectively.

3.4 Conclusions

ESI-MS is widely used to detect non-covalent PL complexes *in vitro* and is increasingly used to quantify these interactions. Underlying the application of the direct ESI-MS assay is the assumption of uniform response factors for P and PL. The results presented here reveal that solution composition can have a measurable

effect on the relative *RF* values and, thereby, influence the reliability of the affinity data measured by ESI-MS. It is found that the introduction of large, high MW solute causes a decrease in $RF_{P/PL}$ values measured for protein-carbohydrate complexes. This effect, which appears to be sensitive to the difference in surface properties of P and PL, is attributed to a reduction in the number of available surface sites in ESI droplets that contain one or more solute molecules and competition between P and PL for these sites.

3.5 Literature Cited

1. Roldos, V.; Javier, C. F.; Jimenez-Barbero, J. *ChemBioChem* **2011,** *12*, 990-1005.

2. Avci, F. Y.; Kasper, D. L. *Annu. Rev. Immunol.* **2010,** *28*, 107.

3. Varki, A.; Cummings, R.; Esko, J.; Freeze, H.; Hart, G.; Marth, J. In *Essentials of Glycobiology*; Cold Spring Harbor Laboratory Press: New York, **1999**; p 653.

4. Bundle, D. R. *Methods Enzymol.* **1994,** *247*, 288-305.

5. Bundle, D. R.; Eichler, E.; Gidney, M. A. J.; Meldal, M.; Ragauskas, A.; Sigurskjold, B. W.; Sinnott, B.; Watson, D. C.; Yaguchi, M.; Young, N. M. *Biochemistry* **1994,** *33*, 5172.

6. Lundquist, J. J.; Toone, E. J. *Chem. Rev.* **2002,** *102*, 555-578.

7. Larsen, K.; Thygesen, M. B.; Guillaumie, F.; Willats, W. G. T.; Jensen, K. J. *Carbohydr. Res.* **2006,** *341*, 1209-1234.

8. Meyer, B.; Peters, T. *Angew. Chem. Intl. Ed.* **2003,** *42*, 864-890.

9. Loo, J. A. *Mass Spectrom. Rev.* **1997,** *16*, 1-23.

10. Daniel, J. M.; Friess, S. D.; Rajagopalan, S.; Wendt, S.; Zenobi, R. *Int. J. Mass Spectrom.* **2002,** *216*, 1-27.

11. Schug, K. A. *Comb. Chem. High Throughput Screen.* **2007,** *10*, 301-316.

12. Soya, N.; Shoemaker, G. K.; Palcic, M. M.; Klassen, J. S. *Glycobiology* **2009,** *19*, 1224-1234.

13. Shoemaker, G. K.; Soya, N.; Palcic, M. M.; Klassen, J. S. *Glycobiology* **2008,** *18*, 587-592.

14. Kitova, E.; El-Hawiet, A.; Schnier, P.; Klassen, J. *J. Am. Soc. Mass Spectrom.* **2012,** *23*, 431-441.

15. Wang, W.; Kitova, E. N.; Klassen, J. S. *Anal. Chem.* **2003,** *75*, 4945-4955.

16. Kitova, E. N.; Kitov, P. I.; Paszkiewicz, E.; Kim, J.; Mulvey, G. L.; Armstrong, G. D.; Bundle, D. R.; Klassen, J. S. *Glycobiology* **2007,** *17*, 1127-1137.

17. Yanes, O.; Villanueva, J.; Querol, E.; Aviles, F. X. *Nat. Protoc.* **2007,** *2*, 119-130.

18. Jecklin, M. C.; Touboul, D.; Jain, R.; Toole, E. N.; Tallarico, J.; Drueckes, P.; Ramage, P.; Zenobi, R. *Anal. Chem.* **2009,** *81*, 408.

19. Yanes, O.; Villanueva, J.; Querol, E.; Aviles, F. X. *Mol. Cell Proteomics* **2005,** *4*, 1602.

20. El-Hawiet, A.; Kitova, E. N.; Arutyunov, D.; Simpson, D. J.; Szymanski, C. M.; Klassen, J. S. *Anal. Chem.* **2012,** *84*, 3867-3870.

21. Zdanov, A.; Bundle, D. R.; Deng, S. J.; MacKenzie, C. R.; Narang, S. A.; Young, M. N.; Cygler, M. *Proc. Natl. Acad. Sci. U.S.A.* **1994,** *91*, 6423-6427.

22. Ritchie, T. K.; Grinkova, Y. V.; Bayburt, T. H.; Denisov, I. G.; Zolnerciks, J. K.; Atkins, W. M.; Sligar, S. G., Chapter 11 Reconstitution of Membrane Proteins in Phospholipid Bilayer Nanodiscs. In *Methods in Enzymology*, Nejat, D., Ed. Academic Press: 2009; Vol. Volume 464, pp 211-231.

23. Bayburt, T. H.; Grinkova, Y. V.; Sligar, S. G. *Nano Letters* **2002,** *2*, 853-856.

24. Connolly, M. L. *J. Mol. Graph.* **1993,** *11*, 139-141.

25. The Pymol molecular graphics system. http://www.pymol.org.

26. Steinbacher, S.; Baxa, U.; Miller, S.; Weintraub, A.; Seckler, R.; Huber, R. *Proc. Natl. Acad. Sci. U.S.A.* **1996,** *93*, 10584-10588.

27.http://glycam.ccrc.uga.edu/ccrc/carbohydrates/cb_newbuilder.jsp?tool=crystallography&option=ff99:glycam06.

28. Tan, M.; Fang, P.; Chachiyo, T.; Xia, M.; Huang, P.; Fang, Z.; Jiang, W.; Jiang, X. *Virology* **2008,** *382*, 115-123.

29. Tan, M.; Huang, P.; Xia, M.; Fang, P.-A.; Zhong, W.; McNeal, M.; Wei, C.; Jiang, W.; Jiang, X. *Journal of Virology* **2011,** *85*, 753-764.

30. Sifurskjold, B. W.; Altman, E.; Bundle, D. R. *Eur. J. Biochem.* **1991,** *197*, 239-246.

31. Armstrong, J. K.; Wenby, R. B.; Meiselman, H. J.; Fisher, T. C. *Biophys. J.* **2004,** *87*, 4259-4270.

32. Zhang, Y.; Liu, L.; Daneshfar, R.; Kitova, E. N.; Li, C.; Jia, F.; Cairo, C. W.; Klassen, J. S. *Anal. Chem.* **2012,** *84*, 7618-7621.

33. Bayburt, T. H.; Sligar, S. G. *FEBS Letters* **2010,** *584*, 1721-1727.

34. Wang, W.; Kitova, E. N.; Klassen, J. S. *Anal. Chem.* **2003,** *75*, 4945-4955.

35. Sun, J.; Kitova, E. N.; Wang, W.; Klassen, J. S. *Anal. Chem.* **2006,** *78*, 3010-3018.

Chapter 4

Measuring Positive Cooperativity using the Direct ESI-MS Assay. Cholera Toxin B Subunit Homopentamer Binding to GM1 Pentasaccharide*

4.1 Introduction

To carry out their functions, proteins must bind to ligands (e.g. other proteins, carbohydrates, lipids, DNA, metal ions or small molecules). Many proteins, in particular enzymes and multisubunit protein complexes, can bind simultaneously to multiple ligand molecules; this may involve multiple copies of the same ligand (homotropic binding) or different ligands (heterotropic binding). Often, binding of one ligand influences the affinities of other ligands. This phenomenon, which represents an important regulatory mechanism in biological systems, is referred to as cooperative binding [1-5]. Cooperativity can be described quantitatively on the basis of Gibbs energy (ΔG) couplings of binding events at different sites [6-7]. Positive cooperativity occurs when ligand binding at one site increases the affinity (association constant, K_a) and decreases ΔG at another site; negative cooperativity arises when ligand binding reduces the affinity (and increases ΔG) at another site [3]. If the binding sites are independent of one another, binding is referred to as non-cooperative [5, 8-9].

* A version of this chapter has been accepted for publication: Lin, H.; Kitova, E. N.; Klassen, J. S. *J. Am. Soc. Mass. Spectrom.*

Despite the recognized importance of cooperative binding in biological processes, quantifying theses effects is challenging [10-11]. The most direct approach to quantifying cooperativity is based on the microscopic association constants (K_a for binding at a specific site) and the changes in K_a (and ΔG) for ligand binding at a given site when another site is occupied [12-14]. However, there are no generally-applicable methods available for monitoring the occupancies of individual ligand binding sites and, consequently, microscopic K_a values are often not experimentally accessible [15-16]. Therefore, a common approach used to evaluate cooperative binding is to compare the trend in macroscopic K_a values (K_a for stepwise ligand binding) with the trend expected for statistical binding. However, most commonly used binding assays, such as isothermal titration calorimetry (ITC) and spectrophotometric techniques, do not directly provide a quantitative measure of the K_a values (or corresponding thermodynamic parameters) for stepwise ligand binding. Consequently, it is often only possible to establish the K_a values by fitting an assumed binding model to the experimental data. In some cases, simplifying assumptions must be made in order to reduce the number of unknown terms [15, 17].

In recent years, electrospray mass spectrometry (ESI-MS) has proven to be a useful tool to detect and quantify protein-ligand interactions, as well as other non-covalent biological complexes, *in vitro* [18-21]. The direct ESI-MS assay [22], which is based on the detection and quantification of the gas-phase ions of free and ligand-bound protein, has been used to measure affinities for a variety of protein-ligand complexes, and in many instances the K_a values agree well with

affinities measured by other analytical methods [23-26]. An important feature of the ESI-MS assay is the ability to study multiple binding equilibria, including stepwise ligand binding, simultaneously. Given that the relative concentrations of the different ligand-bound protein forms can (in principle) be measured directly and the macroscopic K_a values quantified [22], the assay is well suited for studying cooperative binding. Despite this obvious potential, there are surprisingly few examples where the direct ESI-MS assay was used to probe cooperative interactions. In one of the earliest reported examples, Rogniaux *et al.* deduced cooperative ligand binding to a series of enzymes [27]. More recently, Klassen and coworkers used ESI-MS to demonstrate enhanced carbohydrate substrate binding to a glycosyltransferase in the presence of bound donor [21], while Zenobi and coworkers identified from ESI-MS measurements a new ligand binding site resulting from ligand-induced protein conformational changes [28], and Sharon and coworkers used ESI-MS to elucidate the allosteric mechanism for stepwise binding of ATP to the multisubunit protein complex GroEL [29].

Here, we exploit the direct ESI-MS assay to quantify the stepwise binding of the GM1 pentasaccharide β-D-Gal*p*-(1→3)-β-D-Gal*p*NAc-(1→4)[α-D-Neu5Ac-(2→3)]-β-D-Gal*p*-(1→4)-β-D-Glc*p* (referred to here as GM1os) to the B subunit homopentamer of cholera toxin (CT) and to establish whether ligand binding is cooperative. To our knowledge, this is the first reported example of the application of the direct ESI-MS assay to quantify cooperative ligand binding to a multisubunit protein complex. Cholera toxin, which is a member of the AB_5 class of cytotoxins [30], is composed of a catalytically active A subunit and doughnut-

shaped homopentamer of B subunits (B_5), which is responsible for the recognition of host cell receptors and binds selectively to the GM1 ganglioside (β-D-Gal*p*-(1→3)-β-D-Gal*p*NAc-(1→4)[α-D-Neu5Ac-(2→3)]-β-D-Gal*p*-(1→4)-β-D-Glc*p*-Cer) [31-32]. The structure, kinetics and thermodynamics of the interactions of CT holotoxin and CTB_5 with GM1 ganglioside or the soluble GM1os pentasaccharide have been extensively investigated [14, 31-37]. According to the crystal structure of the (CTB_5 + 5GM1os) complex (PDB id 3CHB), each of the five GM1os binding sites (one per subunit) is made up primarily from a B single subunit, with 18 direct or solvent mediated H-bonds between GM1os and amino acid residues located within the subunit and one H-bond with residue Gly33 from an adjacent subunit [31].

Affinity measurements have been performed for CTB_5 binding to soluble GM1os and to GM1 ganglioside in supported bilayers [38] and in vesicles [39] using a variety of techniques, including ITC [14] and surface plasmon resonance (SPR) spectroscopy [37, 40-41]. The binding data are found to be highly dependent on solution conditions (buffer, ionic strength and pH) and on membrane chemistries. Intrinsic association constants ($K_{a,int}$) of between 10^6 and 10^7 M^{-1} have been reported for GM1os [14, 42] and 10^4 to 10^8 for GM1 ganglioside [37, 42-43]. Wiegandt and coworkers, based on the observation of nonlinear Scatchard plots constructed from equilibrium dialysis data, were the first to report that GM1os binding to CTB_5 exhibits positive cooperativity (cooperativity coefficient of 1.25) [34]. Re-analysis of these same data by Schafer and Thakur, using a model comprising seven independent stepwise association reactions, led to the suggestion of a 2-fold

increase in the affinity for the addition of a second (or subsequent) GM1os to CTB_5 [44]. Schon and Freire, from ITC data and a binding model that assumed that cooperativity would manifest itself only through nearest neighbor interactions, reported a 4-fold increase in affinity when the ligand binds adjacent to a subunit that is already occupied [36]. According to this model binding is enhanced further (by a factor 8) when both nearest neighbours are ligand-bound. More recently, Homans and coworkers analyzed ITC data using a simplified form of the nearest neighbour model (Figure 4.1) and reported affinity enhancement factors of 1.9 and 3.4 for GM1os binding to subunits with one or two ligand-bound nearest neighbours, respectively [14]. The ESI-MS binding data reported here serve as direct and conclusive evidence that GM1os binding to CTB_5 exhibits small, positive cooperativity and support the binding model which is proposed by Homans and coworkers [14].

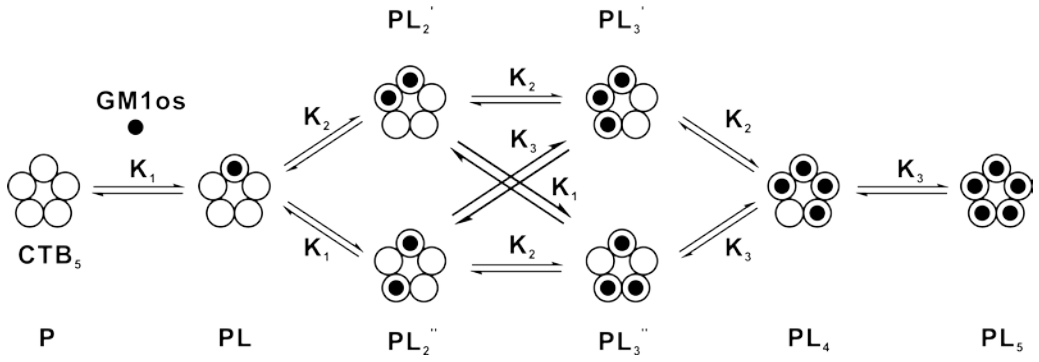


Figure 4.1 Proposed model[14] for the sequential binding of GM1os (L) to the CTB_5 homopentamer (P). Binding is described by three intrinsic association constants, K_1, K_2 and K_3, which represent the case of L binding to a subunit with zero, one or two ligand-bound nearest neighbour subunits, respectively. The

equilibrium concentrations of the eight distinct species (P, PL, $PL_2^{'}$, $PL_2^{''}$, $PL_3^{'}$, $PL_3^{''}$, PL_4 and PL_5) are related through eqs 6a to 6j.

4.2 Experimental

4.2.1 Materials and Methods

Cholera toxin B subunit pentamer (CTB_5, MW 58 020 Da) was purchased from Sigma-Aldrich Canada (Oakville, ON). The protein was concentrated and dialyzed against aqueous 50 mM ammonium acetate and stored at 4°C if not used immediately. The protein concentration was determined using the bicinchoninic acid (BCA) assay calibrated with bovine serum albumin (Pierce, Rockford, USA). The GM1 pentasaccharide β-D-Gal*p*-(1→3)-β-D-Gal*p*NAc-(1→4)[α-D-Neu5Ac-(2→3)]-β-D-Gal*p*-(1→4)-β-D-Glc*p* (GM1os, MW 998.4 Da) was purchased from Elicityl SA (Crolles, France). A stock solution of GM1os was prepared by dissolving a known amount of the solid sample in ultrafiltered water (Milli-Q, Millipore) to yield a concentration of 1 mM. The solution was stored at -20°C until needed.

4.2.2 Mass spectrometry

The binding measurements were carried out with a Bruker 9.4T ApexQe FTICR mass spectrometer (Billerica, MA). Nanoflow ESI (nanoESI) was performed in positive ion mode using borosilicate tubes (1.0 mm o.d., 0.68 mm i.d.), pulled to ~5 μm o.d. at one end using a Sutter Instruments P-2000 micropipette puller

(Novato, CA). The electric field required to spray the solution was established by applying a voltage of ~1.0 kV. The droplets and gaseous ions produced by ESI were introduced into the mass spectrometer through a metal sampling capillary (0.5 mm i.d.). Nitrogen gas at a flow rate of 1.0 L min^{-1} and 90 ^{o}C was used as a drying gas. Both the capillary entrance and exit voltages were held at 280 V. A deflector voltage of 250 V was used. Gaseous ions were transmitted through the first funnel and skimmer held at 150 V and 20 V, respectively, and then through the second funnel and skimmer held at 7.6 V and 5.3 V, respectively. The ions were stored electrodynamically in an rf hexapole for 0.9 s and then accumulated in a hexapole collision cell for 1.2 s. Following accumulation, the ions were transferred into the ion cell for detection. The front and back trapping plates of the cell were maintained at 0.9 and 1.0 V, respectively. The typical base pressure for the instrument was ~1 x 10^{10} mbar. Data acquisition and analysis were performed using ApexControl, version 4.0 (Bruker Daltonics). A minimum of 50 transients with 32k data points per transient were used for each acquisition.

4.2.3 Determination of ligand affinities from ESI-MS data

The general expression for the apparent association constants, $K_{a,q}$, for the stepwise binding of L to P (eq 1) is given by eq 2:

$$P \underset{}{\overset{K_{a,1}}{\rightleftharpoons}} PL \underset{}{\overset{K_{a,2}}{\rightleftharpoons}} PL_2 \rightleftharpoons \ldots \underset{}{\overset{K_{a,q}}{\rightleftharpoons}} PL_q \qquad (1)$$

$$K_{a,q} = \frac{R_q}{R_{q-1}([L]_0 - \frac{(R_1 + 2R_2 + \ldots qR_q)[P]_0}{1 + R_1 + R_2 + \ldots R_q})} \qquad (2)$$

where $[L]_0$ and $[P]_0$ are the initial concentration of L and P, respectively, and the R_q terms represent the concentration ratios of ligand-bound to free protein, i.e., $[PL_q]/[P]$. Provided that the ionization and detection efficiencies (i.e., response factors) of the ions corresponding to the ligand-bound and free protein species are equivalent, the R_q terms can be determined from the abundances (*Ab*) of the gaseous ligand-bound and free P ions, summed over all charge states (n), eq 3:

$$R_q = \frac{Ab(PL_q)}{Ab(P)} = \frac{\sum_n Ab(PL_q)^{n+}}{\sum_n Ab(P)^{n+}} = \frac{[PL_q]}{[P]} \quad (3)$$

For a protein with *Q* equivalent and independent binding sites, the apparent $K_{a,q}$ values are related by statistical factors that reflect the number of occupied and unoccupied binding sites, eq 4:

$$K_{a,q} / K_{a,q-1} = (q - 1)(Q - q + 1)/(q(Q - q + 2)) \quad (4)$$

Intrinsic association constant $K_{a,int}$ can be found from any of the $K_{a,q}$ values using the general expression:

$$K_{a,int} = qK_{a,q} /(Q - q + 1) \quad (5)$$

If the ligand binding sites are not equivalent or in cases of cooperative binding (positive or negative), the relationship given by eq 5 will not be valid, *vide infra*.

4.3 Results and discussion

Measurements were carried out on aqueous ammonium acetate (30 mM, pH 6.9) solutions of CTB_5 (8.5 µM) and GM1os at concentrations ranging from 0 to 55 µM. Shown in Figure 4.2 are representative mass spectra acquired in the absence

of GM1os and presence of GM1os (at concentrations of 6 μM, 17.5 μM and 50 μM).

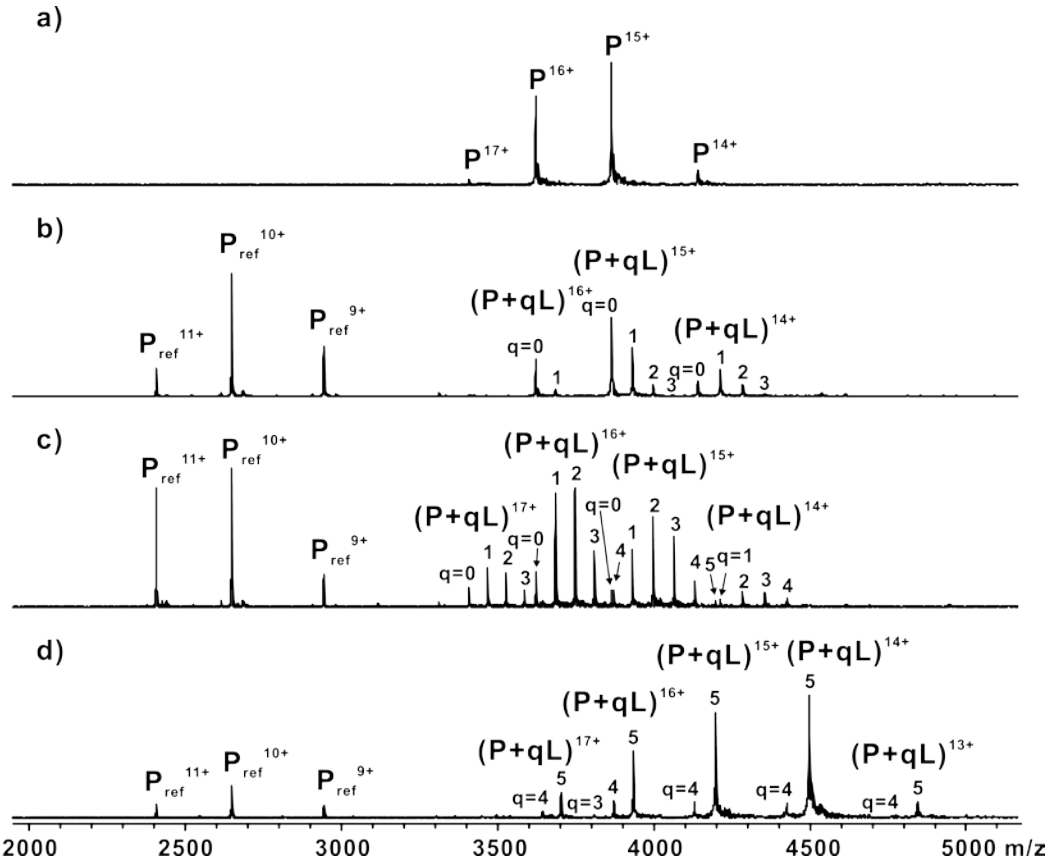


Figure 4.2 ESI mass spectra acquired for aqueous ammonium acetate (30 mM) solutions (22 °C at pH 6.9) of CTB_5 (8.5 μM) and varying concentrations of GM1os (a) 0 μM (b) 6 μM (c) 17.5 μM and (d) 50 μM. A reference protein (4.5 μM) was added into the solution to identify the occurrence of nonspecific ligand binding. The number of molecules of GM1os bound to CTB_5 is indicated by q.

In the absence of GM1os, the only protein ions detected correspond to protonated homopentamer, $(CTB_5 + nH)^{n+} \equiv CTB_5^{n+}$ at charge states n = 14 – 17. At low GM1os concentrations (6 μM), ions corresponding to CTB_5 bound to 1, 2

or 3 GM1os are detected, i.e., $(CTB_5 + qGM1os)^{n+}$ ions where q = 1 - 3, while at the higher concentrations investigated $(CTB_5 + 4GM1os)^{n+}$ and $(CTB_5 + 5GM1os)^{n+}$ ions were also detected. It should be noted that a reference protein (P_{ref}) was also added to the solutions in order to monitor the occurrence of nonspecific ligand-protein binding during the ESI process [45]. However, there was no evidence of nonspecific binding in any of the measurements.

Apparent $K_{a,q}$ values for the stepwise addition of GM1os to CTB_5 were determined from the ESI mass spectra as described in the Experimental section. Listed in Table 4.1 are the $K_{a,q}$ values measured at twelve different GM1os concentrations and the corresponding average values.

Table 4.1. Apparent association constants ($K_{a,q}$) for the stepwise binding of GM1os to CTB_5 measured at 22 °C and pH 6.9 by ESI-MS.[a,b]

[GM1os] (μM)	$K_{a,1}$ (10^6 M^{-1})	$K_{a,2}$ (10^6 M^{-1})	$K_{a,3}$ (10^6 M^{-1})	$K_{a,4}$ (10^6 M^{-1})	$K_{a,5}$ (10^6 M^{-1})
6.0	15.6 ± 0.7	6.6 ± 0.4	3.0 ± 0.1	-	-
7.5	13.8 ± 0.6	5.8 ± 0.2	2.9 ± 0.1	-	-
10.0	14.1 ± 1.1	6.5 ± 0.8	3.6 ± 0.8	-	-
12.5	14.4 ± 0.6	5.9 ± 0.5	3.5 ± 0.1	3.4 ± 0.1	-
15.0	16.2 ± 0.7	7.1 ± 0.6	4.1 ± 0.3	2.6 ± 0.2	-
17.5	16.7 ± 0.9	7.5 ± 0.9	3.8 ± 0.8	2.3 ± 0.4	3.2 ± 0.2
22.5	15.6 ± 0.4	8.7 ± 0.6	5.2 ± 0.2	2.8 ± 0.7	1.1 ± 0.5
27.5	16.0 ± 0.6	8.7 ± 0.5	4.8 ± 0.2	2.8 ± 1.1	1.3 ± 0.2
35.0	-	9.1 ± 0.7	3.2 ± 0.2	3.8 ± 0.9	1.7 ± 0.6
45.0	-	-	-	3.7 ± 0.6	1.6 ± 0.1
50.0	-	-	-	3.0 ± 0.2	1.9 ± 0.2
55.0	-	-	-	-	2.0 ± 0.1
Average	16.0 ± 1.2	7.3 ± 1.3	3.8 ± 0.8	3.1 ± 0.5	1.9 ± 0.6
Calculated[c]	16.0 ± 1.2	6.4 ± 0.5	3.2 ± 0.2	1.6 ± 0.1	0.6 ± 0.05

a. CTB_5 (8.5 μM) in 30 mM ammonium acetate.
b. Errors correspond to one standard deviation.
c. Calculated, assuming that the five binding sites are equivalent and independent, apparent association constants based on the $K_{a,int}$ value determined by ESI-MS.

Inspection of the values in Table 4.1 reveals that the apparent $K_{a,q}$ values decrease with increasing q. This effect is expected due to the reduction in the number of available binding sites. However, the actual reduction in $K_{a,q}$ is less than expected in the case where the binding sites are equivalent and independent (Table 4.1). Assuming that the five binding sites are equivalent, $K_{a,int}$, which corresponds to $(1/5)K_{a,1}$, is found to be $(3.2 \pm 0.2) \times 10^6$ M^{-1}. This value is slightly smaller (by a factor of 2) than reported by Homans and coworkers ($(6.4 \pm 0.3) \times 10^6$ M^{-1}), although the solution conditions used in their study, most notably pH, were slightly different [14]. From $K_{a,int}$, the $K_{a,q}$ values expected in the absence of cooperative binding can be calculated: $K_{a,2} = (4/2)K_{a,int} = (6.4 \pm 0.5) \times 10^6$ M^{-1}, $K_{a,3} = (3/3)K_{a,int} = (3.2 \pm 0.2) \times 10^6$ M^{-1}. $K_{a,4} = (2/4)K_{a,int} = (1.6 \pm 0.1) \times 10^6$ M^{-1}, $K_{a,5} = (1/5)K_{a,int} = (0.6 \pm 0.05) \times 10^5$ M^{-1}. It can be seen that the measured and calculated values do not agree, the measured $K_{a,q}$ values being larger. This comparison establishes conclusively that GM1os binding to CTB_5 exhibits positive cooperativity, in agreement with previous proposals [14, 44]. The ESI-MS derived $K_{a,q}$ values also provide a means of directly quantifying the magnitude of this effect.

In order to quantify the magnitude of the cooperative binding effects, the $K_{a,q}$ values were analyzed using the model proposed by Homans and coworkers (Figure 4.1) [14]. According to this model, stepwise ligand binding can be described using three intrinsic association constants, K_1, K_2 and K_3, which represent the case of ligand binding to a subunit with zero, one or two ligand-bound nearest neighbours, respectively. It follows that the equilibrium concentrations of the

eight distinct species present in solution (referred to here as P, PL, $PL_2^{'}$, $PL_2^{''}$, $PL_3^{'}$, $PL_3^{''}$, PL_4 and PL_5) are related by eqs 6a – 6j:

$$5K_1 = \frac{[PL]}{[P][L]} \tag{6a}$$

$$K_2 = \frac{[PL_2^{'}]}{[PL][L]} \tag{6b}$$

$$K_2 = \frac{[PL_3^{'}]}{[PL_2^{'}][L]} \tag{6c}$$

$$K_2 = \frac{[PL_4]}{[PL_3^{'}][L]} \tag{6d}$$

$$K_3 = \frac{[PL_4]}{[PL_3^{''}][L]} \tag{6e}$$

$$\frac{K_3}{5} = \frac{[PL_5]}{[PL_4][L]} \tag{6f}$$

$$K_1 = \frac{[PL_2^{''}]}{[PL][L]} \tag{6g}$$

$$K_2 = \frac{[PL_3^{''}]}{[PL_2^{''}][L]} \tag{6h}$$

$$K_1 = \frac{[PL_3^{''}]}{[PL_2^{'}][L]} \tag{6i}$$

$$K_3 = \frac{[PL_3^{'}]}{[PL_2^{''}][L]} \tag{6j}$$

It should be noted that the $PL_2^{'}$ and $PL_2^{''}$ species, as well as the $PL_3^{'}$ and $PL_3^{''}$ species can't be distinguished by ESI-MS because they have identical molecular

weights. It follows that K_1 and K_3 can be calculated directly from $K_{a,1}$ and $K_{a,5}$, eqs 7 and 8:

$$K_{a,1} = 5K_1 = \frac{[PL]}{[P][L]} \quad (7)$$

$$K_{a,5} = \frac{K_3}{5} = \frac{[PL_5]}{[PL_4][L]} \quad (8)$$

while K_2 is related to K_1 and K_3 by eq 9:

$$K_1K_2K_3=K_2^3 \quad (9)$$

Using this approach, the values of K_1, K_2 and K_3 were determined to be $(3.2 \pm 0.2) \times 10^6$ M^{-1}, $(5.5 \pm 1.2) \times 10^6$ M^{-1} and $(9.5 \pm 3.5) \times 10^6$ M^{-1}, respectively. According to this analysis, binding is enhanced by a factor of 1.7 when one of the neighbouring subunits is already bound to GM1os and a factor of 2.9 when both neighbouring subunits are ligand-bound. These values are in reasonable agreement with enhancement factors of 1.9 and 3.4, inferred from ITC data, reported by Homans and coworkers [14].

Although the similarities in the enhancement factors obtained from the ESI-MS and ITC data offer support for the proposed binding model (Figure 4.1), they do not, on their own, conclusively establish that the model correctly describes the experimental data. To further test the appropriateness of this model, the K_1, K_2 and K_3 values extracted from the ESI-MS data were used to predict the corresponding $K_{a,2}$, $K_{a,3}$ and $K_{a,4}$ values. It can be seen that the predicted $K_{a,q}$ values ($K_{a,2} = K_1 + K_2 = (8.7 \pm 1.2) \times 10^6$ M^{-1}; $K_{a,3} = (K_2 + K_3)/(K_2/K_1 + 1) = (5.5 \pm 1.6) \times 10^6$ M^{-1}; $K_{a,4} = K_3/(K_3/K_2 + 1) = (3.4 \pm 1.6) \times 10^6$ M^{-1}) are in good agreement with the values obtained from ESI-MS measurements. These findings

confirm that the proposed binding model properly describes the stepwise binding of GM1os to CTB_5.

4.4 Conclusions

In summary, apparent $K_{a,q}$ values were measured for the stepwise binding of GM1os to CTB_5 at pH 6.9 and 22°C using the direct ESI-MS assay. Analysis of the binding data provides direct evidence that GM1os binding exhibits small, positive cooperativity. The ESI-MS data are consistent with a binding model, proposed by Homans and coworkers, in which ligand binding to CTB_5 can be described by three intrinsic K_a values that represent the case of ligand binding to a subunit with zero, one or two ligand-bound nearest neighbours (Figure 4.1). According to the ESI-MS results, binding to a subunit located next to a single ligand-bound subunit results in an affinity enhanced by a factor of 1.7, while binding is enhanced by a factor of 2.9 when both nearest neighbors are bound to ligand. These results highlight the strength of the direct ESI-MS assay for studying stepwise ligand binding to proteins and quantifying cooperativity effects. This study also lays necessary groundwork for the development of a new ESI-MS technique, employing competitive ligand binding, to quantify the interactions between CTB_5 and lipid bilayer supported GM1 ganglioside.

4.5 Literature cited

1. Changeux, J.-P., *Cold Spring Harb. Symp. Quant. Biol.*, **1963**, 28, 497-504.

2. Changeux, J.-P. and M.M. Rubin, *Biochem.*, **1968**, 7, 553-560.

3. Levitzki, A. and D.E. Koshland, *Proc. Natl. Acad. Sci. U.S.A.*, **1969**, 62, 1121-1128.

4. Koshland, D.E. and K.E. Neet, *Annu. Rev. Biochem.*, **1968**, 37, 359-410.

5. Whitty, A., *Nat. Chem. Biol.*, **2008**, 4, 435-439.

6. Ackers, G.K., M.A. Shea, and F.R. Smith, *J. Mol. Biol.*, **1983**, 170, 223-242.

7. Forsén, S. and S. Linse, *Trends in Biochemical Sciences*, **1995**, 20, 495-497.

8. Deng, L., E. Kitova, and J. Klassen, *J. Am. Soc. Mass. Spectrom.*, **2013**, 24, 49-56.

9. Mammen, M., S.K. Choi, and G.M. Whitesides, *Angew. Chem., Int. Ed.*, **1998**, 37, 2754-2794.

10. Gabler, E., *Pharmazie*, **1977**, 32, 739-747.

11. Edelstein, S.J. and N. Le Novère, *J. Mol. Biol.*, **2013**, 425, 1424-1432.

12. Linse, S. and W.J. Chazin, *Protein Sci.*, **1995**, 4.

13. Tochtrop, G.P., et al., *Proc. Natl. Acad. Sci. U.S.A.*, **2002**, 99, 1847-1852.

14. Turnbull, W.B., B.L. Precious, and S.W. Homans, *J. Am. Chem. Soc.*, **2004**, 126, 1047-1054.

15. Onufriev, A. and G.M. Ullmann, *J. Phys. Chem. B*, **2004**, 108, 11157-11169.

16. Di Cera, E., Thermodynamic Theory of Site-Specific Binding Processes in Biological Macromolecules. 1995, Cambridge, U.K.: Cambridge Univ. Press.

17. Garcés, J., L. Acerenza, E. Mizraji, and F. Mas, *J Biol Phys*, **2008**, 34, 213-235.

18. Loo, J.A., *Mass Spectrom. Rev.*, **1997**, 16, 1-23.

19. Daniel, J.M., S.D. Friess, S. Rajagopalan, S. Wendt, and R. Zenobi, *Int. J. Mass Spectrom.*, **2002**, 216, 1-27.

20. Soya, N., G.K. Shoemaker, M.M. Palcic, and J.S. Klassen, *Glycobiology*, **2009**, 19, 1224-1234.

21. Shoemaker, G.K., N. Soya, M.M. Palcic, and J.S. Klassen, *Glycobiology*, **2008**, 18, 587-592.

22. Kitova, E., A. El-Hawiet, P. Schnier, and J. Klassen, *J. Am. Soc. Mass. Spectrom.*, **2012**, 23, 431-441.

23. Jørgensen, T.J.D., P. Roepstorff, and A.J.R. Heck, *Anal. Chem.*, **1998**, 70, 4427-4432.

24. El-Hawiet, A., et al., *Anal. Chem.*, **2012**, 84, 3867-3870.

25. El-Hawiet, A., E. Kitova, L. Liu, and J. Klassen, *J. Am. Soc. Mass. Spectrom.*, **2010**, 21, 1893-1899.

26. Liu, L., E. Kitova, and J. Klassen, *J. Am. Soc. Mass. Spectrom.*, **2011**, 22, 310-318.

27. Rogniaux, H., et al., *Anal. Biochem.*, **2001**, 291, 48-61.

28. Jecklin, M., D. Touboul, C. Bovet, A. Wortmann, and R. Zenobi, *J. Am. Soc. Mass. Spectrom.*, **2008**, 19, 332-343.

29. Dyachenko, A., R. Gruber, L. Shimon, A. Horovitz, and M. Sharon, *Proc. Natl. Acad. Sci. U.S.A.*, **2013**, 110, 7235-7239.

30. Odumosu, O., D. Nicholas, H. Yano, and W. Langridge, *Toxins*, **2010**, 2, 1612-1645.

31. Merritt, E.A., et al., *J. Protein Sci.*, **1994**, 3, 166-175.

32. Heyningen, S.V., *Science (New York, N.Y.)*, **1974**, 183, 656-7.

33. Fishman, P.H., J. Moss, and J.C. Osborne, *Biochem.*, **1978**, 17, 711-716.

34. Sattler, J., G. Schwarzmann, J. Staerk, W. Ziegler, and H. Wiegandt, *Hoppe-Seyler's Z. Physiol. Chem.*, **1978**, 6, 719-723.

35. Miller, I.R., H. Vinkler, and E. Yavin, *Bioelectrochemistry and Bioenergetics*, **1989**, 22, 365-377.

36. Schön, A. and E. Freire, *Biochem.*, **1989**, 28, 5019-5024.

37. Kuziemko, G.M., M. Stroh, and R.C. Stevens, *Biochem.*, **1996**, 35, 6375-6384.

38. Moran-Mirabal, J.M., et al., *Biophysical Journal*, **2005**, 89, 296-305.

39. Williams, T.L. and A.T.A. Jenkins, *J. Am. Chem. Soc.*, **2008**, 130, 6438-6443.

40. MacKenzie, C.R., T. Hirama, K.K. Lee, E. Altman, and N.M. Young, *J. Biol. Chem.*, **1997**, 272, 5533-5538.

41. Borch, J., F. Torta, S.G. Sligar, and P. Roepstorff, *Anal. Chem.*, **2008**, 80, 6245-6252.

42. Lauer, S., B. Goldstein, R.L. Nolan, and J.P. Nolan, *Biochem.*, **2002**, 41, 1742-1751.

43. Baksh, M.M., M. Jaros, and J.T. Groves, *Nature*, **2004**, 427, 139-141.

44. Schafer, D.E. and A.K. Thakur, *Cell Biophys.*, **1982**, 4, 25-40.

45. Sun, J., E.N. Kitova, W. Wang, and J.S. Klassen, *Anal. Chem.*, **2006**, 78, 3010-3018.

Chapter 5

Conclusion and Future Work

This work describes the application of ESI-MS methods to study the non-covalent protein-carbohydrate interactions. The first two research projects focus on phenomena associated with the ESI process and their influence on the application of ESI-MS to study protein conformation and ligand binding affinities. The last research project highlighted the potential of ESI-MS for quantifying cooperative ligand binding to multiple subunits protein complexes.

In Chapter 2, new evidence for the occurrence of rapid, electrostatic-induced unfolding of acidic proteins in negatively charged ESI droplets was reported and investigated. The extent of unfolding was monitored by changes in the CSD and the *ACS*, was found to be sensitive to the concentration of the solution "buffer", NH_4OAc. For neutral solutions of a highly acidic sub-fragment (B3C) of the C-terminal carbohydrate-binding repeat region of the exotoxin, toxin B (TcdB), at high *I*, >80 mM, the mass spectra exhibit a relatively narrow CSD and constant *ACS*, consistent with the protein having a compact structure. However, for solutions at lower *I*, the proteins exhibit a much broader CSD and a substantially larger (absolute) *ACS*, consistent with unfolding of the protein. In contrast, the CSD and *ACS* measured in positive ion mode are essentially independent of *I* (over the range investigated) and consistent with a folded protein. The results of proton nuclear magnetic resonance (NMR) and circular dichroism (CD) spectroscopy measurements and gel filtration chromatography (GFC)

suggest that protein B3C existed predominantly in a folded state in neutral aqueous solutions with $I > 10$ mM. Control experiments which were performed on B3C in a series of solutions with high I at pH 5 to 9 rule out the possibility that structural changes are related to ESI-induced pH changes. It is proposed that the unfolding of B3C, observed in negative mode for solutions with low I, occurs in the ESI droplets and arises due to Coulombic repulsion between the negatively charged residues of the protein and droplet surface charge. The results of ESI-MS measurements performed on the mutants B4A and B4B, which contain fewer acidic residues than B3C, also reveal a shift to higher absolute *ACS* at low I. However, in both cases the magnitude of the change is smaller than observed for B3C, consistent with the proposed electrostatic-induced unfolding mechanism.

In Chapter 3, the deleterious effects of large and high MW solute (polymers and non-covalent assemblies) on protein-ligand affinity measurements carried out using the direct ESI-MS assay were investigated. The presence of high MW solute, that does not interact with the protein (P) or ligand (L) of interest, is shown to have a measurable effect on the relative response factor (*RF*) values and, thereby, influence the reliability of the affinity measured for protein-carbohydrate complexes by ESI-MS. The magnitude of the effect, which appears to be sensitive to the differences in the surface properties of P and PL, is attributed to a reduction in the number of available surface sites in ESI droplets that contain one or more solute molecules and competition between P and PL for these sites. A similar effect which was achieved upon introduction of a small surfactant to the solution provides qualitative support for this hypothesis.

In Chapter 4, direct ESI-MS assay was applied to measure apparent association constants for the stepwise addition of one to five GM1os to CTB_5 at pH 6.9 and 22°C. The intrinsic association constant ($K_{a,int}$), which was established from the apparent association constant for the addition of a single GM1os to CTB_5, was found to be $(3.2 \pm 0.2) \times 10^6\ M^{-1}$. The result is in reasonable agreement with the reported value of $(6.4 \pm 0.3) \times 10^6\ M^{-1}$, which was measured using isothermal titration calorimetry at pH 7.4 and 25 °C[1]. Analysis of the binding data also provides direct and unambiguous evidence that GM1os binding exhibits small, positive cooperativity. It was found that the binding between GM1os and CTB_5 is sensitive to the number of ligand-bound nearest neighbour subunits in CTB_5, with the affinities enhanced by a factor of 1.7 and 2.9 when binding occurs next to one or two ligand-bound subunits, respectively. These findings provide quantitative support for the binding model proposed by Homans and coworkers[1], More importantly, it lays necessary groundwork for the development of a new ESI-MS technique, employing competitive ligand binding, to quantify the interactions between CTB_5 and lipid bilayer supported GM1 ganglioside.

There are several possible extensions of the current studies. As mentioned in Chapter 3, the presence of high MW solute, that does not interact with the P or L of interest, is shown to have a measurable effect on the relative *RF* values, which is a big challenge in application of indirect ESI-MS assay[2] to carry out quantitative study of carbohydrate interactions with viral protein particles.[3] After understanding the origin of this differential effect on the *R* values of P and PL individually, we can carry out more measurements titrating in different molecules

which don't interact with P or L of interest to find out the suitable candidate that may have the opposite effect (an increase) on *RF* values. Detailed calculation of SAS area can be assigned to single molecule to help fish out possible candidate. In this way, the deleterious effect on *RF* values can be corrected to broaden the application of ESI-MS assay (direct or indirect).

Another possible extension of the current work is to investigate the interactions between CTB_5 and its Glycosphingolipids (GSL) receptors. We have already measured the $K_{a,int}$ and cooperativity in the stepwise binding model of GM1os to CTB_5. And recently, the use of nanodiscs (NDs), has emerged as a promising and versatile technology for studying soluble protein interactions with membrane bound receptors, such as GSL, in a biologically-relevant lipid milieu.[4-5] Combining the utility of NDs and ESI-MS assay, we can then determine the microscopic association constants for interactions between CTB_5 and its membrane bound GSL receptors such as GM1 by using a new binding assay, *proxy ligand* ESI-MS assay. This assay may also reveal new insights into the phenomenon of GSL clustering and its effects on protein binding.[6-7]

Shown in Figure 5.1 is the illustrative experimental scheme of *proxy ligand* ESI-MS assay. It is based on the use of a suitable L_{proxy} ($L_{proxy} \equiv L$) one that binds specifically to P with a known affinity, and for which the interaction can be quantified using the direct ESI-MS assay. The fraction of L_{proxy} (GM1os) bound to P (CTB_5), which is determined directly from the ESI spectrum, is sensitive to the fraction of the other ligand of interest, L_{ND} (NDs containing GM1) in solution.

The binding of CTB_5 to GM1 can then be quantified based on the changes, in the relative abundance ratios of CTB_5-GM1os complex to free CTB_5.

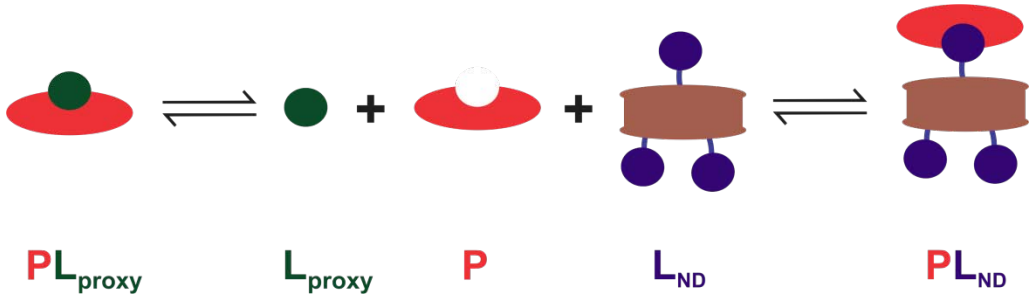


Figure 5.1 Schematic representation of *proxy ligand* ESI-MS assay in quantifying CTB_5 and its membrane bound GSL receptors GM1 ganglioside interactions.

The potential strategy is to make different NDs containing varying amount of GM1 (between 1 and 10%) and those NDs can be titrated into solution samples containing known amount initial of CTB_5 and GM1os. Based on the well studied stepwise binding model of CTB_5 and GM1os with small cooperativity, different initial amount of protein ($[P]_0$) and proxy ligand ($[L_{proxy}]_0$) can be mixed together to achieve different starting equilibrium point. Finally, the fitting of the different experimental titration curves to the theoretical ones based on the determined binding model can offer quantitative results for interactions between CTB_5 and its glycosphingolipid receptor GM1.

5. 1 Literature cited

1. Turnbull, W.B., B.L. Precious, and S.W. Homans, *J. Am. Chem. Soc.*, **2004**, 126, 1047-1054.

2. El-Hawiet, A., et al., *Anal. Chem.*, **2012**, 84, 3867-3870.

3. Tan, M., et al., *Virology*, **2008**, 382, 115-123.

4. Nath, A., W.M. Atkins, and S.G. Sligar, *Biochemistry*, **2007**, 46, 2059-2069.

5. Borch, J., F. Torta, S.G. Sligar, and P. Roepstorff, *Anal. Chem.*, **2008**, 80, 6245-6252.

6. Shi, J., et al., *J. Am. Chem. Soc.*, **2007**, 129, 5954-5961.

7. Lingwood, D., et al., *Nat. Chem. Biol.*, **2011**, 7, 260-262.