University of Alberta

Study of Non-Covalent Protein-Carbohydrate Interactions using Electrospray Ionization Mass Spectrometry

by

Amr El-Hawiet

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

Doctor of Philosophy

Department of Chemistry

Fall 2012
Edmonton, Alberta

To my parents

To my wife and daughters

Abstract

This thesis describes the development and application of electrospray ionization mass spectrometry (ESI-MS) methods to study protein-carbohydrate interactions in vitro. The affinities (K_a) of two recombinant fragments of *Clostridium difficile* toxins (TcdA & TcdB) and a library of the most abundant human milk oligosaccharides (HMOs) were quantified using the direct ESI-MS assay. The results of the study revealed that both of the toxin fragments recognize, albeit weakly, a variety of HMOs ranging in size from tri- to heptasaccharides. The results of molecular docking simulations suggest that a disaccharide moiety (lactose or lactosamine) is the core HMO recognition element for both toxin fragments. The protective effects of HMOs fractions, extracted from human milk, were tested using the verocytotoxicity neutralization assay. However, the results revealed that the HMOs fractions do not significantly inhibit the cytotoxic effects of TcdA or TcdB.

Combining the direct ESI-MS assay and competitive binding, two new ESI-MS assays were developed. The r*eference ligand* ESI-MS method allows for the quantification of protein-ligand complexes that are prone to dissociation in the gas phase while the *proxy protein* ESI-MS method allows for the quantification of carbohydrate binding to large protein complexes that cannot be directly detected by ESI-MS. Using the *reference ligand* ESI-MS method, affinities of two carbohydrate-binding proteins for monosaccharide ligands were quantified, while the *proxy protein* ESI-MS method was used to quantify the interactions of tail spike protein of bacteriophage P22 (180 kDa) together with its mutant to their

natural receptors. The results of binding measurements performed using these new methods were in excellent agreement with the reported values.

A catch-and-release (CaR) ESI-MS assay for screening carbohydrate libraries against target proteins was also developed. Ligands with moderate affinity (10^4 - 10^6 M^{-1}) were successfully detected from mixtures containing >200 carbohydrates. Additionally, the absolute affinities were estimated from the abundance of free and ligand-bound protein ions determined from the ESI mass spectrum. Multiple low affinity ligands (~10^3 M^{-1}) were successfully detected in mixtures containing >20 carbohydrates. The use of ion mobility separation, performed on deprotonated carbohydrate ions following their release from the complex, allowed for the positive identification of isomeric ligands.

Table of Contents

List of Tables

7 determined by the direct ESI-MS assay performed on the individual ligands and a library consisting of 23 carbohydrates.

List of Figures

Figure 5.15 **(**a) Representative ESI mass spectrum obtained for a solution of TcdB-B3 (P, 10 μM), αLA (P_{ref}, 15 μM) and 23 carbohydrates (10 μM each) acquired in positive ion mode, where **a** = $(P+\mathbf{L36})^{+11}$, **b** = $(P+\mathbf{L12})^{+11}$, **c** = $(P+\mathbf{L10})^{+11}$, **d** = $(P+\mathbf{L134})^{+11}$, **e** = $(P+\mathbf{L40})^{+11}$, **f** = $(P+\mathbf{L7})^{+11}$, **g** = $(P+\mathbf{L47})^{+11}$, h = $(P+\mathbf{L8})^{+11}$, **i** = $(P+\mathbf{L6})^{+11}$, while **b′** = $(P_{ref}+\mathbf{L12})^{+7}$, **c′** = $(P_{ref}+\mathbf{L10})^{+7}$, **d′** = $(P_{ref}+\mathbf{L134})^{+7}$, **f′** = $(P_{ref}+\mathbf{L7})^{+7}$, **g′** = $(P_{ref}+\mathbf{L47})^{+7}$, **h′** = $(P_{ref}+\mathbf{L8})^{+7}$, **i′** = $(P_{ref}+\mathbf{L6})^{+7}$, **j′** = $(P_{ref}+\mathbf{L137})^{+7}$. 176

List of Abbreviations

Ab	Abundance
AIMS	Aspiration ion mobility spectrometry
ATD	Arrival time distribution
CaR	Catch and release
CID	Collision induced dissociation
Con A	Concanavalin A
CRM	Charge residue model
CVFF	Consistent valence force field
Da	Dalton
DC	Direct current
DESI	Desorption electrospray ionization
DMS	Differential-mobility spectrometry
DTIMS	Drift-time ion mobility spectrometry
ELISA	Enzyme-linked immunosorbent assay
ESI	Electrospray ionization
ESI-MS	Electrospray ionization mass spectrometry
Fab	Antigen binding fragment
FAC-MS	Frontal affinity chromatography-mass spectrometry
FAIMS	Field-asymmetric waveform ion mobility spectrometry
FT-ICR-MS	Fourier transform ion cyclotron resonance mass spectrometry
FWHM	Full width at half maximum

H-bond	Hydrogen bond
H/D exchange	Hydrogen/deuterium exchange
HMOs	Human milk oligosaccharides
IEM	Ion evaporation model
IMS	Ion mobility separation
IS	Internal standard
ITC	Isothermal titration microcalorimetry
K_a	Association constant
$K_{a,app}$,	Apparent association constant
L	Ligand
L_{ref}	Reference ligand
Lyz	Lysozyme
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MW	Molecular weight
m/z	Mass-to-charge ratio
nanoESI	Nanoflow electrospray ionization
NMR	Nuclear magnetic resonance
P	Protein
P_{proxy}	Proxy protein
P_{NS}	Non interacting protein
P_{ref}	Reference protein
QTOF	Quadrupole time-of-flight

R	Abundance ratio
RF	Response factor
scFv	Single chain variable fragment
S/N	Signal-to-noise ratio
SPR	Surface plasmon resonance
SRIG	Stacked ring ion guide
T-wave	Travelling wave
TcdA	*Clostridium difficile* toxin A
TcdB	*Clostridium difficile* toxin B
TOF	Time of flight
TSP	Tailspike protein
TWIMS	Travelling voltage wave ion mobility mass spectrometry
Ubi	Ubiquitin
VL	Virtual ligand
WT	Wildtype
αLA	α-lactalbumin
ΔH	Solvation enthalpy
ΔG	Gibbs free energy
ΔS	Entropy

Chapter 1

Study of Non-Covalent Protein-Carbohydrate Interactions using Electrospray Ionization Mass Spectrometry*

1.1 Introduction

Carbohydrates are the most abundant biological molecules.[1] They are commonly found on the surface of cells, in the form of glycopeptides, glycoproteins and glycolipids and can bind to suitable protein receptors (e.g. lectins, antibodies and carbohydrate-processing enzymes) in solution or on cell surfaces.[2] Carbohydrate-protein interactions play critical roles in a wide range of physiological and pathological cell functions, such as inflammation, cell-cell and cell-matrix interactions, signal transduction, fertility, development and cancer metastasis.[3] They also represent prerequisite first steps for the infection of hosts by many microbes, including viruses, bacteria and their toxins, parasites and fungi, and are implicated in subsequent immune responses.[3,4] The association of carbohydrate-protein complexes is driven primarily by the formation of hydrogen bond (H-bond) networks and van der Waals contacts.[5] Solvent effects also strongly influence the thermodynamics of carbohydrate-protein binding.[4] In their unbound form, carbohydrate ligands, as well as the protein residues involved in their binding, are typically well-solvated in an aqueous environment. Complete or partial dehydration of the binding partners, which is necessary for complex formation, is

* Portions of this chapter has been published: Kitova, E. N.; El-Hawiet, A.; Paul D. Schnier, P. D.; Klassen, J. S. *J. Am. Soc. Mass Spectrom*. **2012,** *23*, 431-441 and El-Hawiet, A.; Kitova, E. N.; Klassen, J. S. *Biochemistry* **2012,** *51*, 4244-4253.

energetically costly and, as a result, carbohydrate-protein complexes often (but not always) exhibit low association constants (K_a), in the ~10^3 M^{-1} range.[6] To overcome the low affinities typical of individual carbohydrate-protein interactions, many carbohydrate-binding proteins possess multiple carbohydrate binding sites and exploit multivalent binding to achieve high avidities.[7]

The detection of carbohydrate-protein interactions and their characterization (structure and thermodynamic and kinetic parameters) is both of fundamental importance and facilitates the design of carbohydrate-based therapeutics to treat a variety of diseases and infections. There are a number of established analytical methods to identify and quantify carbohydrate-protein interactions in vitro, each with particular strengths and weaknesses. Isothermal titration microcalorimetry (ITC) is generally considered the "gold standard" technique for quantifying the thermodynamic parameters of complex formation, and is the only assay that directly provides a measure of the enthalpy of association. Conventional ITC instruments suffer from low sensitivity and generally require large amounts (~mg) of pure protein and ligand. However, new ITC technologies, such as the Nano ITC™, have improved sensitivity and substantially lower sample requirements.[8]

Surface plasmon resonance (SPR) spectroscopy represents one of the most widely used assays for evaluating the affinities of carbohydrate-protein interactions.[9,10] The technique also enables the direct determination of association and dissociation rate constants. 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

binding interaction. Indeed, there are examples where the ITC and SPR spectroscopy yield divergent binding data for the same ligand-protein interaction.[11,12]

Frontal affinity chromatography combined with mass spectrometry detection (FAC–MS) allows for the analysis of mixtures of compounds for specific protein interactions and the determination of the corresponding K_a values.[13] The method involves the continuous infusion of ligands through a column wherein the protein target is immobilized on a solid support. The ligands, which are detected by electrospray ionization (ESI) MS, are eluted according to their binding affinities for the target protein, thereby enabling the relative affinities to be easily established. A limitation of the FAC-MS assay is the requirement for immobilization of the target protein, which is impractical in some cases and may affect the binding properties of the protein.[13]

Enzyme linked immunosorbent assay (ELISA) is another widely used method for quantifying carbohydrate-protein interactions.[14] While there are many ways of implementing ELISA, the assay requires the immobilization of one of the binding partners, which is incubated with solutions containing the other binding partner, often in the presence of a soluble inhibitor/competitive binder. The ELISA method, once set-up, is fast and relatively sensitive. However, the assay is quite labour intensive and often requires conjugation of the ligand, which can limit its applicability.

Glycan (carbohydrate) microarrays have become a popular tool for the discovery of carbohydrate interactions with proteins and protein complexes. The arrays consist of oligosaccharides that are attached to a solid support, usually through a covalent linker.[15,16] Protein targets are incubated with the array and, following a washing step,

specific interactions are identified, usually using a fluorescence-based readout.[17] A dense presentation of the oligosaccharides serves to mimic the situation encountered on cell surfaces that allows for multivalent interactions with low affinity carbohydrate ligands.[18] However, the technique can provide, at best, semi-quantitative binding data.

Recently, ESI-MS analysis has earned a place among the arsenal of tools available for identifying and quantifying (stoichiometry and affinity) carbohydrate-protein interactions, as well as other ligand-protein complexes. The ESI-MS measurements can be categorized as either "direct" or "indirect" in nature. The direct ESI-MS assay relies on the direct detection of the gas phase ions of free and ligand-bound protein. One of the earliest examples of direct ESI-MS detection of specific ligand-protein complexes in aqueous solution was reported in 1991 by Ganem, Li and Henion and involved interaction of lysozyme with N-acetylglucosamine substrates.[19] Soon after the first reports of the successful detection of non-covalent ligand-protein complexes by ESI-MS, the potential of the technique for measuring the affinities of ligand-protein interactions began to be exploited.[20,22-39] 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 the B_5 homopentamer of the Shiga-like toxin type I.[20] The K_a values measured by ESI-MS were found to be in a good agreement with values measured by ITC.[21] 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).[27-39]

The direct ESI-MS assay possesses a number of strengths, including its simplicity (no labeling or immobilization of protein or ligand is required), speed (individual K_a measurements can usually be completed within a few minutes). Additionally, when performed using nanoflow ESI, which operates at solution flow rates in the 10-100 nL min^{-1} range, the assay normally consumes pmol or less of analyte per analysis.

Two important features of the ESI-MS assay are the abilities to directly establish the stoichiometry of protein complexes and to measure multiple binding equilibria simultaneously. These features enable the determination of both the macroscopic and microscopic K_a values for sequential binding of L to P. As a result, ESI-MS is ideally suited for characterizing allosteric binding. The ESI-MS assay also naturally lends itself to monitoring and quantifying protein-ligand interactions in solutions containing mixtures of ligands and/or proteins.[40-48] Direct ESI-MS measurements of ligand-protein affinities are normally carried out at ambient temperature. However, with the development of temperature-controlled ESI devices, it is also possible to evaluate K_a over a range of solution temperatures. The magnitude of the corresponding enthalpy and entropy of association (ΔH_a and ΔS_a, respectively) can be estimated from a van't Hoff analysis of the temperature-dependence of the K_a values.[49-52]

However, the ESI-MS assay, like all binding assays, has limitations. The method relies on being able to detect and accurately quantify the free and ligand-bound protein ions, which is not always possible. But, the ESI-MS assay is also well-suited for competitive binding experiments in which multiple proteins or ligands compete for binding partners. The combination of competitive binding experiments and ESI-MS

detection, which relies on direct ESI-MS analysis of protein-ligand complexes to deduce the strength of other protein-ligand interactions in the solution, has been exploited in numerous studies to extract binding data that could not be measured directly by ESI-MS.

1.2 Direct ESI-MS binding assay

The assay is based on the detection and quantification of free and ligand-bound protein ions by ESI-MS. The K_a for a given protein-ligand interaction is determined from the ratio (R) of total abundance (Ab) of all ligand-bound and free protein ions, as measured by ESI-MS for solutions of known initial concentrations of protein ($[P]_o$) and ligand ($[L]_o$). For a 1:1 protein-ligand complex (eq 1.1), K_a is calculated using eq 1.2: [39]

$$P + L \rightleftharpoons PL \tag{1.1}$$

$$K_a = \frac{R}{[L]_o - \frac{R}{1+R}[P]_o} \tag{1.2}$$

where R is given by eq 1.3:

$$\frac{[PL]_{eq}}{[P]_{eq}} = \frac{Ab(PL)}{Ab(P)} = R \tag{1.3}$$

Normally, K_a for a particular protein-ligand interaction is not determined at a single concentration of P and the L but rather from measurements performed at a number of different concentrations or from a titration experiment, wherein the concentration of one analyte (normally P) is fixed and the concentration of the other is

varied.[25] The value of K_a can be extracted using nonlinear regression analysis of the experimentally determined concentration-dependence of the fraction of ligand-bound protein, i.e., $R/(R+1)$, which is given by the following expression:

$$R/(R+1) = \frac{1 + K_a[P]_o + K_a[L]_o - \sqrt{(1 + K_a[P]_o - K_a[L]_o)^2 + 4K_a[L]_o}}{2K_a[P]_o} \quad (1.4)$$

It follows that K_a values accessible with the direct ESI-MS binding assay range from approximately 10^3 to 10^7 M^{-1} which suits the study of the protein-carbohydrate interactions. However, interactions with much larger K_a values can be probed using competitive binding and direct ESI-MS measurements, *vide infra*.

1.3 Potential pitfalls of direct ESI-MS assay

Like all the other previously mentioned analytical methods, the direct ESI-MS has its own limitations. Any physical or chemical process that alters the equilibrium abundance ratio of bound-to-free protein during the ESI process and in the gas phase from that present in bulk solution will lead to incorrect K_a values and, potentially, obscure the true binding stoichiometry. There are four common sources of error associated with the ESI-MS measurements: *i*) non-uniform response factors, *ii*) in-source dissociation, *iii*) nonspecific ligand-protein binding and *iv*) ESI-induced changes in solution pH and temperature. Each of these sources of error is briefly described below, along with current strategies for minimizing their effects on the binding measurements.

1.3.1 Non-uniform response factors

As described above, the abundances of P and PL measured by ESI-MS are related to the solution concentration by a response factor (*RF*), which collectively accounts for the ionization and detection efficiencies, eq 1.5:

$$[PL]/[P] = RF_P Ab(PL)/RF_{PL} Ab(P) = RF_{P/PL}(Ab(PL)/Ab(P)) \quad (1.5)$$

Underlying eqs 1.2 and 1.4 is the assumption of uniform *RF* values (i.e., $RF_{P/PL} \approx 1$). This assumption is generally valid in cases where L is small compared to P, such that the size and surface properties of the P and PL are similar.[22-26,32,34,37-39,53-55] While there are no firm guidelines 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., $MW_{PL}/MW_P \leq 110\%$.[37] However, there are cases where the ESI-MS response of a protein complex is significantly different than the response of the free protein.[56,57] It is important to note that *RF*s depend on many factors - the size and structure of P and PL, the ESI conditions and the instrumental parameters used for the measurements.

A variety of strategies have been developed to minimize the effects of nonuniform *RF*s on the determination of K_a values. One approach involves the introduction of the $RF_{P/PL}$ term as an adjustable parameter in an appropriate binding model, which is fit to the experimental data.[56-61] 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.[59] Furthermore, this approach is based on the assumption 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 that is similar (MW and surface activity) to the analyte of interest, but which does not bind to L.[61] The 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. An alternative strategy involves monitoring the abundance of L, relative to that of an IS, as $[P]_o$ is varied.[62] 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 $[P]_o$.

1.3.2 In-source dissociation

Collision-induced dissociation of gaseous ions of PL complexes during ESI-MS analysis can alter the relative abundance of PL and P ions. Ion-source dissociation is caused by the collisional heating of gaseous ions which may occur at various stages during the ion sampling process, such as within the heated metal sampling capillary (if used), in the nozzle (or orifice)-skimmer region, and during accumulation of ions within external rf multipole storage devices (e.g., hexapole).[31,39,63-65] In-source dissociation will necessarily decrease the magnitude of K_a and in the extreme case, where no PL ions survive to detection, in-source dissociation results in a false negative. The influence of in-source dissociation on binding measurements depends on the configuration of the ion source used, the choice of instrumental parameters and the size and gas-phase stability of the complex. The stability of PL complexes in the gas phase is determined, in part, by the nature of the specific interactions in solution. Complexes stabilized in solution predominantly by weak or a small number of intermolecular interactions generally exhibit low gas phase stabilities and are

susceptible to in-source dissociation.[27,63,66,67] However, it is important to note that the gas-phase stabilities of PL complexes generally do not parallel the solution affinities. For example, some PL complexes, which are stabilized by strong ionic interactions in solution, exhibit low gas phase stabilities,[63] while some PL complexes formed by hydrophobic bonding are quite stable in the gas phase.[26,64,65] Usually, the occurrence of in-source dissociation can be identified from changes in *R* resulting from changes in ion source parameters, in particular voltage differences in regions of high pressure (e.g. nozzle-skimmer voltages), that influence the internal energy of the ions. Recently, the use of pulsed hydrogen-deuterium exchange (HDX) for identifying the occurrence of in-source dissociation involving multiprotein complexes was demonstrated.[68]

Low temperatures (drying gas, sampling capillary), low potentials across lens elements, and short accumulation times are essential for obtaining reliable K_a values for PL complexes susceptible to in-source dissociation. However, there are usually trade-offs between the use of so-called "gentle" source conditions and signal intensity. Thus, a balance must be found between minimizing the extent of in-source dissociation and achieving adequate protein ion signal. In cases where gentle sampling conditions do not eliminate the occurrence of in-source dissociation, the employment of stabilizing additives may prove beneficial. For example, the addition of imidazole to solution, at high concentration (>1 mM), has been shown to prevent gas phase dissociation of the ions of a number of different PL interactions, including protein-carbohydrate, protein-fatty acid and protein-small molecule complexes.[26,63,65] 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.[63] Additionally, the use of imidazole, which has a relatively high gas phase basicity and a relatively low gas phase acidity,[69,70] may also lead to a reduction in the charge states of the protein complex ions. The lower charge state complex ions may exhibit higher kinetic stabilities and be more resistant to in-source dissociation.[71] Recently, it was shown that the introduction of imidazole vapor to the ion source also protects complexes against in-source dissociation.[65] It has also been shown that the presence of a high partial pressure of SF_6 in the ion source reduces the extent of in-source dissociation for some complexes.[65]

1.3.3 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. 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 PL complexes can be understood in the context of the charge residue model (CRM) [72] of ESI (see section 1.5).

According to the CRM, the initial ESI droplets undergo solvent evaporation followed by fission, releasing several small multiply charged nanodroplets. 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. The probability of the nanodroplets containing more than one analyte molecule increases with analyte concentration.[73] Therefore, a general strategy

for minimizing the occurrence of nonspecific ligand binding involves limiting the concentration of L. However, high ligand concentrations (>0.05 mM) are typically required to detect weak (K_a <10^4 M^{-1}) PL interactions. In such cases, nonspecific binding is often unavoidable

A number of strategies have been proposed to correct ESI mass spectra for the occurrence of nonspecific ligand binding.[68,74-81] The most direct approach is the *reference protein* method, which involves the addition of a non-interacting reference protein (P_{ref}) to solution to quantitatively correct for nonspecific ligand binding to the protein/complexes of interest.[74] The method is based on the assumption that nonspecific ligand binding is random, 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 assumption that, in a given ESI-MS experiment, the distribution of nonspecifically bound L is independent of the nature of the protein has been rigorously tested and shown to be valid for a variety of "ligands", including neutral and charged carbohydrates, amino acids, peptides and divalent metal ions.[74-76,82] It follows that the "true" abundance of a given PL_q species ($Ab(PL_q)$) 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:

$$Ab(PL_q) = [Ab_{app}(PL_q) - f_{1,Pref}Ab(PL_{q-1}) - f_{2,Pref}Ab(PL_{q-2}) - \cdots f_{q,Pref}Ab(P)]/f_{0,Pref} \qquad (1.6)$$

where $f_{q,Pref}$ is the fractional abundance of P_{ref} bound to q molecules of L. This correction method has been successfully used in binding studies performed on a

variety of protein-ligand interactions, including protein-carbohydrate and protein-metal ion complexes.[38,82]

An alternative method, called the *reporter molecule* method, was developed to identify the occurrence of nonspecific protein-protein binding during the ESI process.[80] To implement the method, a non-interacting small molecule (M_{rep}) is added to the solution, at elevated concentration. Differences in the distributions of the small molecule bound nonspecifically to the different protein species present (e.g. monomer versus dimer) is used to establish the occurrence of nonspecific protein-protein binding. Another approach named the *nonspecific probe* method was recently developed to identify the occurrence of nonspecific binding between small molecules during the ESI process.[81] In this method, a non-interacting protein (P_{NS}) is added to solution and the distributions of small molecules bound nonspecifically to P_{NS} is used to establish whether small molecule complexes originated from solution or they were formed by nonspecific binding. A weakness with both the *reporter molecule* and *nonspecific probe* methods is that they do not allow ESI mass spectra to be quantitatively corrected for the occurrence of nonspecific binding.

1.3.4 ESI-induced changes in solution pH and temperature

The K_a values for protein-ligand interactions in aqueous solution are generally sensitive to pH and temperature. Both the pH and the temperature of the solution may be altered by the ESI-MS measurement and lead to changes in K_a, particularly when low solution flow rates are used. Electrochemical reactions, which occur at the electrode in the ESI tip, can alter the composition of the solution.[83] In aqueous solution the dominant electrochemical reactions occurring at a chemically inert electrode are

oxidation (positive ion mode) and reduction (negative ion mode) of H_2O leading to the production of H_3O^+ and OH^-, respectively. At low solution flow rates (<100 nL/min), the resulting pH changes can be large, >1 pH unit after 30 min of spraying.[39] The use of ESI solutions with a high buffer capacity or short spraying times (<10 min) is sufficient to minimize errors in K_a introduced by pH changes. One must also be on guard against inadvertent changes to the temperature of the solution. Most commercial ESI sources rely on heating of the droplets to accelerate/assist with the desolvation of ions to improve sensitivity. This is commonly achieved by applying heated air or N_2 as a drying gas in the region of the ESI tip or sampling of the ESI droplets into a heated metal capillary. Exposure of the ESI tip to heated gas or having the tip in proximity of a heated metal capillary can lead to changes in temperature of a few degrees or more, particularly when using low solution flow rates.

1.3.5 Other challenges

In addition to the sources of error outlined above, there are also technical issues that currently limit the utility of the direct ESI-MS assay. Among these is the general incompatibility of the assay with "physiological" buffers. In protein-ligand binding studies, the primary role of the buffer is to keep the protein stable and minimize protein aggregation. For many proteins there is often a narrow range of concentration, pH, and ionic strengths, which provides a binding-competent, active protein. Mass spectrometric studies often employ aqueous ammonium acetate solutions (1-200 mM), with minimal nonvolatile salts or detergents added to the solution. A variety of strategies have been proposed for ESI-MS analysis of solutions containing physiological buffers at relevant concentrations, including the use of high ammonium

acetate concentrations,[84,85] carrying out ESI in the presence of high velocity air (gas),[86] and decoupling the sample solution from the ESI process through the use of desorption electrospray ionization (DESI).[87]

1.4 Indirect ESI-MS assays

While many carbohydrate-protein interactions can reliably be studied using the direct ESI-MS assay, the technique has limitations. As noted above, it is not generally applicable to interactions for which the corresponding gaseous ions readily dissociate at room temperature. Ligand binding to very large or heterogeneous proteins or proteins complexes also poses a significant challenge due to instrumental limitations (mass range and resolution). Additionally, the technique is generally incompatible with the analysis of interactions involving membrane proteins or insoluble cellular receptors. In cases where direct ESI-MS analysis is not feasible, it is sometimes possible to quantify the interactions of interest by combining the direct ESI-MS assay with competitive protein or ligand binding. Several such strategies have been developed, for example, indirect ESI-MS methods have been proposed in this thesis to quantify, labile carbohydrate-protein interactions that dissociate during ESI-MS analysis *"reference ligand* method", and carbohydrate interactions with large proteins and protein complexes, which cannot be directly analyzed "*proxy protein* method". The complete description of these methods is given in chapters 3 and 4, respectively.

Several other indirect ESI-MS assays have been developed. For example, Konermann and coworkers have shown that ligand-protein interactions can be quantified based on changes in the diffusion coefficient of a ligand in the presence and absence of protein, as measured by ESI-MS.[66] An alternative strategy for detecting

specific ligand-protein interactions and measuring their affinities involves the quantification of the concentration of free ligand in solution. For example, Leary and co-workers demonstrated such an approach, which is based on ESI-MS analysis of a library of ligands before and after incubation with the immobilized enzyme, for identifying potential inhibitors.[42] Another strategy, which is suitable for the analysis of high affinity ligand-protein interactions, relies on the use of a L_{ref}, with known affinity for the protein of interest, and the determination of the relative abundance of the ligand and L_{ref} by ESI-MS.[88] A related method uses the changes of the abundance ratio of ligands upon changes in protein concentration to establish relative affinities.[89]

1.5 The present work

The work described in this thesis focuses on the application of the direct ESI-MS assay to quantify protein-carbohydrate interactions and the development of new methodologies to expand the versatility and utility of ESI-MS for quantifying protein-ligand interactions.

The work descriped in chapter 2 focuses on examining quantitatively the binding of recombinant fragments of the C-terminal cell-binding domains of the two large exotoxins, toxin A (TcdA) and toxin B (TcdB), expressed by *Clostridium difficile* and a carbohydrate library consisting of the most abundant neutral and acidic human milk oligosaccharides (HMOs) using the direct ESI-MS assay. The study shows that both toxins fragments investigated bind specifically, but weakly, to HMOs ranging in size from tri- to heptasaccharides. Notably, five of the HMOs tested bind to both toxins. The results of molecular docking simulations, taken together with the experimental binding data, provided more insight into the interaction modalities of HMOs with the

toxin fragments. The results of a Verocytotoxicity neutralization assay reveal that the HMOs fractions, extracted from human milk, do not significantly inhibit the cytotoxic effects of TcdA or TcdB, which is attributed to the very weak intrinsic affinities that the toxins exhibit towards the HMOs.

Chapter 3 describes an ESI-MS approach, the *reference ligand* ESI-MS method, for quantifying protein-ligand interactions that are prone to in-source dissociation and, therefore, difficult to detect directly. This method is based on the direct ESI-MS assay and competitive ligand binding. It involves the addition of a reference ligand (L_{ref}) to the ESI solution, which binds specifically to the protein, at the same binding site as the ligand of interest, with known affinity and forms a stable protein-ligand complex in the gas phase. The fraction of the protein bound to L_{ref}, which is determined directly from the ESI mass spectrum, is sensitive to the fraction of the protein bound to ligand of interest in solution and enables the affinity of the protein for ligand of interest to be determined.

Chapter 4 proposes an ESI-MS method for quantifying protein-ligand complexes that are not readily directly detected by ESI-MS. The *proxy protein* ESI-MS method combines direct ESI-MS binding measurements with competitive protein-ligand binding. It involves the use of a proxy protein (P_{proxy}), which interacts specifically with the ligand of interest with known affinity and can be detected directly by ESI-MS, to quantitatively monitor the extent of ligand binding to the protein of interest. A modified form of the *proxy protein* ESI-MS method was applied to account for real-time changes in ligand concentration due to the enzymatic hydrolyzing activity of the protein of interest.

A catch and release (CaR-ESI-MS) assay for screening carbohydrate libraries against target proteins is reported in chapter 5. This method was applied to three protein systems. Ligands with moderate affinity (10^4 - 10^6 M^{-1}) were successfully detected from mixtures containing >200 carbohydrates. The absolute affinities of the binders were estimated from the abundance of free and ligand-bound protein ions determined from the ESI mass spectrum. Multiple low affinity ligands (~10^3 M^{-1}) were successfully detected in mixtures containing >20 carbohydrates. The use of ion mobility separation, performed on deprotonated carbohydrate ions following their release from the complex, allowed for the positive identification of isomeric ligands.

The experimental work described in this thesis was performed using nano ESI on a quadrupole-ion mobility separation-time-of-flight (Q-IMS-TOF) mass spectrometer and Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometers. The fundamental aspects of the main experimental techniques used in this research will be reviewed in the following section.

1.6 Electrospray ionization

The mechanism of ESI[90,91] 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.

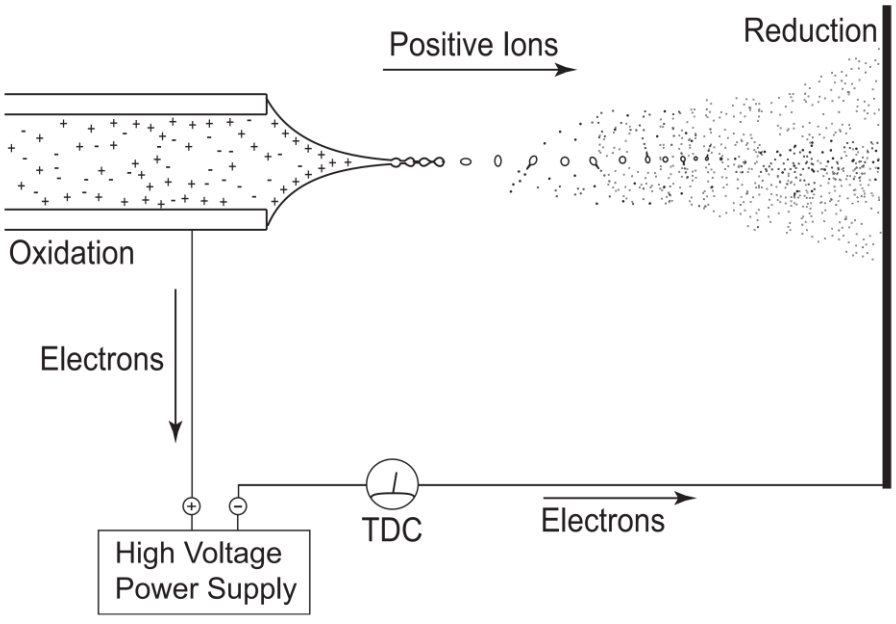


Figure 1.1* Illustration of major processes of ESI in the positive ion mode

* Modified from reference 90

(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.[92] 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 and a fine jet emerges from the cone tip. The surface of the jet, which is charged by an excess of positive ions, will break up into small charged droplets due to the repulsion between the charges on the jet.

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

The charged droplets produced at the spray needle will shrink due to solvent evaporation causing an increase in the electric field normal to the surface of the droplet 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 droplet gets smaller the repulsion between the charges at the surface overcomes the cohesive force of the surface tension causing a coulomb fission of the droplet, also called a coulomb explosion.[90] This leads to fission of the droplet that typically releases a jet of small, charged progeny droplets.

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

Two mechanisms have been proposed to account for the formation of gas-phase ions from the very small and highly charged droplets.

i) The ion evaporation model (IEM):[93] This model, proposed by Iribarne and Thomson, 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 is experimentally well-supported for small (in)organic ions.

ii) The charged residue model (CRM):[72] It was proposed by Dole and coworkers for high molecular mass analytes (e.g. proteins). This model assumes that the charged macro-ions are produced from very small droplets, which contain one macromolecule. As this droplet evaporates completely the charges on the droplet are transferred to the macromolecule.

Recently Michael Gross and co-workers[94] proposes 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.

1.7 MS instrumentation

1.7.1 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 (Figure 1.2).

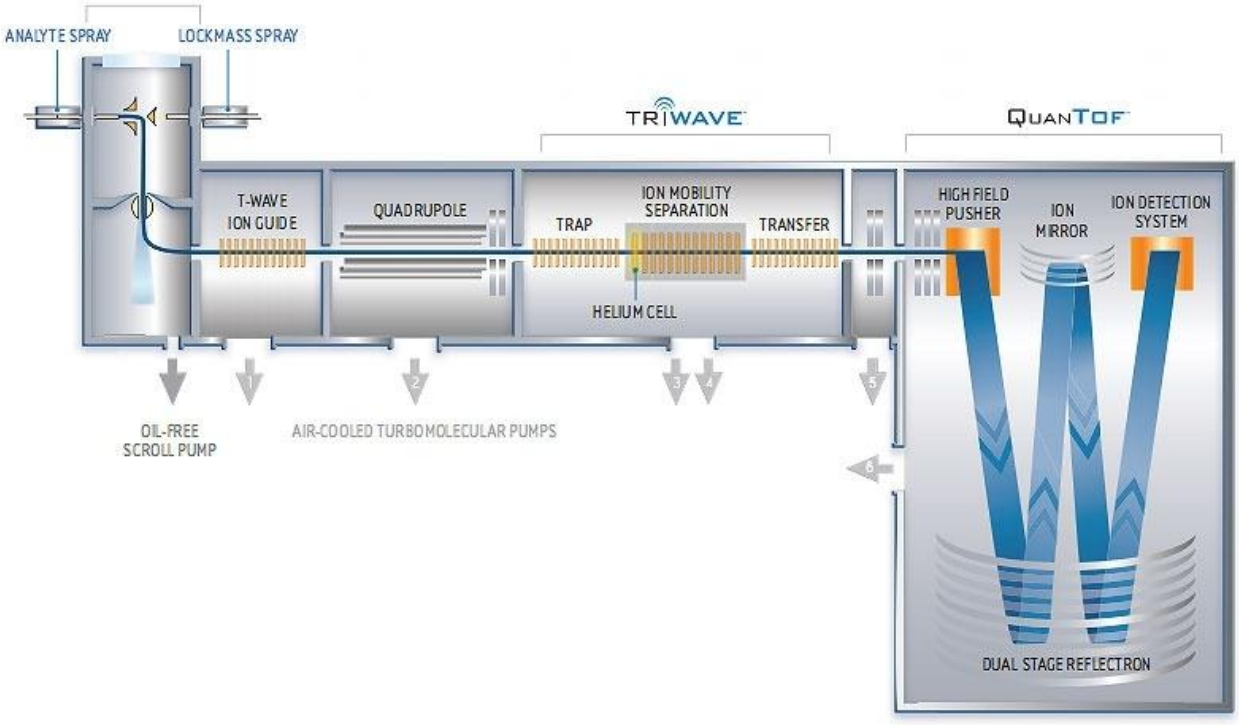


Figure 1.2 A schematic diagram of the Synapt HDMS system*

Gaseous ions are produced by nanoESI using borosilicate capillaries (1.0 mm o.d., 0.68 mm i.d.) pulled to ~5 µm. A platinum wire is inserted into the nanoESI tip and a capillary voltage of 1.0-1.3 kV is applied to carry out ESI in positive ion mode. The cone voltage is set to 50-90 V and the source block temperature is maintained at 70-90 ºC. Ions are transmitted through a quadrupole mass filter to the ion mobility section of the instrument (Triwave), containing three traveling wave (T-Wave) ion guides, the TRAP T-Wave traps and accumulates ions, after which the ions are released in a packet into the IMS T-Wave, where ion mobility separation is performed. A high-pressure helium-filled cell at the front of the IMS T-Wave cell is used to minimize scattering and/or fragmentation as ions are introduced into the high N_2 pressure region. The TRANSFER T-Wave delivers the mobility separated ions to an

* http://www.waters.com

orthogonal acceleration (oa) -TOF mass analyzer (QuanTof™) equipped with a high field pusher and a dual-stage reflectron.

1.7.1.1 Quadrupole Mass Analyzers

Quadrupoles are four cylindrical metal rods that are accurately poitioned 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.[95] 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; 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.

1.7.1.2 Time of Flight (TOF) Mass Analyzers

For the TOF analyzers, the physical property that is measured during an analysis is the flight time of the ions.[95-99] Mass-to-charge (m/z) ratios 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 equation 1.7:

$$(m/z)^{1/2} = t(\frac{\sqrt{2eV_S}}{L}) \qquad (1.7)$$

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^2, the terms in parentheses being constant. This equation demonstrates that 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 analyzers suffer from the drawback that ions of the same m/z can reach the detector at different times, due to initial energy distribution, resulting in peak broadening and poor resolution. The reflectron TOF analyzer compensates for this energy distribution 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 more 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.

Coupling TOF analyzers with continuous ionization techniques (like ESI) requires the use of the orthogonal acceleration (oa) technique. Ion optics focuses the ions resulting from ESI into a parallel beam and directs it to the orthogonal accelerator. A pusher is then pulsed to introduce ions into the orthogonally situated flight tube. During the time that the ions continue their flight in the flight tube, the orthogonal accelerator is refilled with new ion beam.

1.7.1.3 Ion mobility MS

Ion mobility (IM) spectrometry is a gas-phase electrophoretic technique that allows analytes to be distinguished on the basis of their mass, charge and collision cross section (i.e., size and shape).[100-105] In IM spectrometry, a combination of gas flow and electrical fields are used to move the ions towards the drift region. A drift gas is present in the drift region at a constant pressure. An ion passing through the buffer gas experiences a number of collisions, which impede its progress towards the detector. Larger ions with greater collision cross sections experience more collisions than smaller ions and therefore take longer to traverse the drift tube.[100] The mobility of an ion (K) is determined by equation 1.8:

$$K = \frac{d}{t_d E} \tag{1.8}$$

where t_d is the time taken to traverse the drift cell, d is the length of the drift cell and E is the electric field gradient. There are five methods of ion mobility separation that are currently used in mass spectrometry, drift-time ion mobility spectrometry (DTIMS),[106] aspiration ion mobility spectrometry (AIMS),[107] differential-mobility spectrometry (DMS),[108] which is also called field-asymmetric waveform ion mobility spectrometry (FAIMS), trapped ion mobility spectrometry (TIMS) and traveling-wave ion mobility spectrometry (TWIMS).[109] The detail of TWIMS, which is used in this thesis, is outlined briefly as follows:

The TWIMS operates with a high field which is applied to one segment of the cell and swept sequentially through the cell one segment at a time in the direction of ion migration, separating ions based on their mobility.[103] The IM section of (the Synapt

HDMS system-Waters Corp., Milford, USA) comprises three travelling wave-enabled stacked ring ion guides (SRIGs) as shown in more detail in Fig. 1.2.

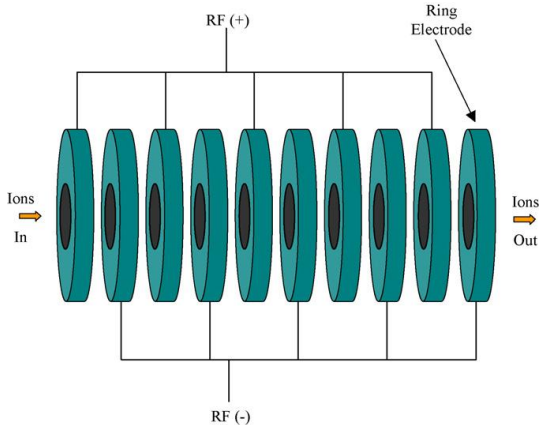


Figure 1.3 A stacked ring ion guide (SRIG).

The traveling wave stacked rings ion guide comprises a series of planar electrodes arranged orthogonally to the ion transmission axis, as shown in Fig.1.3. Opposite phases of a RF voltage are applied to adjacent electrodes and provide a radially confining effective potential barrier. To propel ions through the gas, a transient DC voltage is superimposed on the RF applied to a pair of adjacent electrodes in a repeating sequence along the length of the device. The series of potential hills generated are subsequently applied to the next pair of electrodes downstream at regular time intervals providing a continuous sequence of 'travelling waves'. The ions within the device are driven away from the potential hills and consequently are carried through the device with the waves, minimising their transit time. Ion species of high mobility slip behind the waves less often than species of low mobility and so are transported through the device more quickly, thus mobility-based separation of ions occurs.[105] A particular advantage of the TWIMS device over most drift tubes is that through use of ion accumulation and radial ion confinement, the sensitivity of the mass

spectrometer is not compromised when operating in mobility mode and allowing investigations on analytically significant levels of sample.[105]

1.7.2 Fourier Transform Ion Cyclotron Resonance (FT-ICR) Mass Spectrometers.

Shown in Figure 1.4a is a schematic diagram of the Bruker Apex-II-FTICR MS used in the present work. Gaseous ions are produced by nanoESI performed at atmospheric pressure on buffered aqueous solutions containing analyte molecule 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.

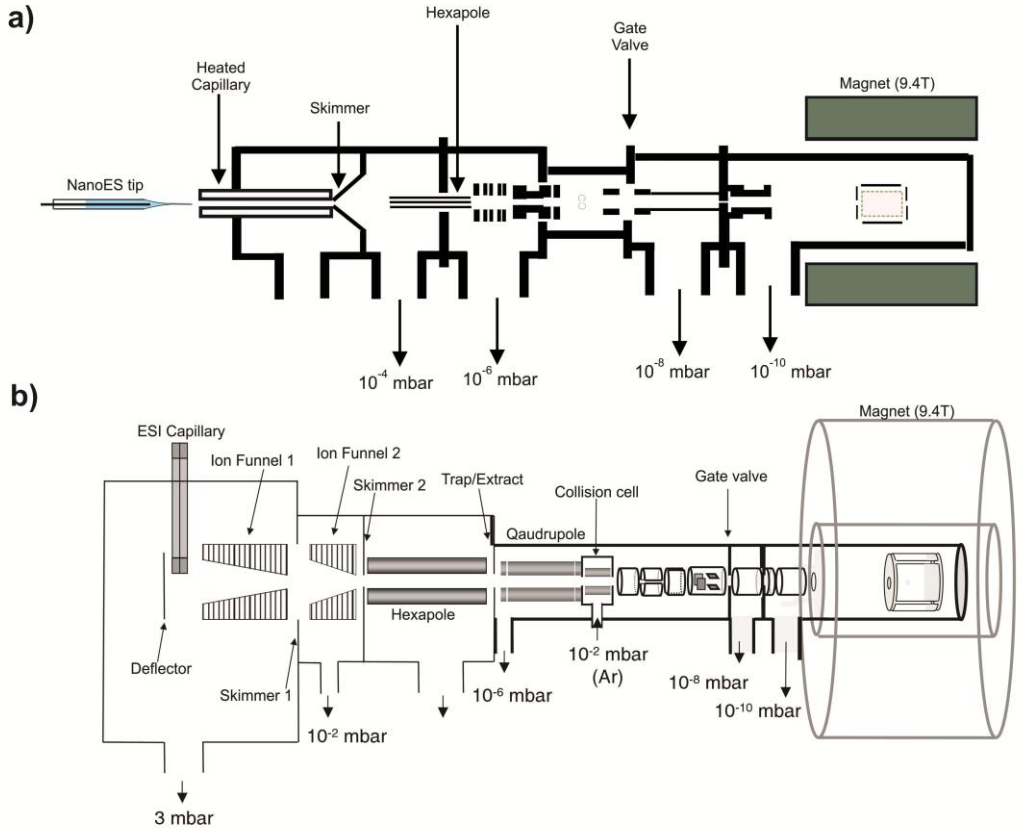


Figure 1.4 Schematic diagrams of (a) the Bruker Apex-II 9.4T ESI-FTICR-MS and (b) the Bruker Apex-Qe 9.4T ESI-Qh-FTICR-MS.

The other FT-ICR mass spectrometer used in this thesis is a Bruker Apex-Qe ESI-Qh-FTICR-MS (Figure 1.4b). Apex-Qe is a hybrid quadrupole/FTICR mass spectrometer, in which two mass analyzers are combined. The first quadrupole acts as a mass filter to select and isolate targeted analyte efficiently for tandem MS (MS/MS) analysis. In the present work, however, the quadrupole was operated in radio frequency (RF)-only mode as it acted as a wide band-pass filter to transmit ions to 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 orthogonally, with the assistance of a deflector, into the first funnel stage. They are focused into a defined ion beam, with the help of DC and RF voltages, to pass through the aperture of skimmer 1. The ions then enter the second funnel leading to skimmer 2 in the next vacuum stage at ~0.1 mbar. They are decelerated before they experience further cooling in the next vacuum stage of the source hexapole, where ions can be accumulated for a defined time or simply passed through, before they enter the Qh-interface.

1.7.2.1 FT-ICR

The principle of the FT-ICR mass analyzer is based on ion cyclotron motion, which arises from the interaction of an ion trapped within a spatially homogeneous static magnetic field.[110-113] Figure 1.5 illustrates the cyclotron motion of a positive ion subjected to a static magnetic field, B, directed into the plane of the paper. The cyclotron frequency, ω_c, is described in equation 1.9:

$$\omega_c = \frac{zeB}{m} \tag{1.9}$$

where z is the charge on the ion, e is the elementary charge, 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 1.9 is that all ions of a given m/z rotate at the same frequency, independent of their velocities. The ultrahigh resolution achieved by FT-ICR MS is a direct result of insensitivity of the cyclotron frequency to the kinetic energy of an ion.

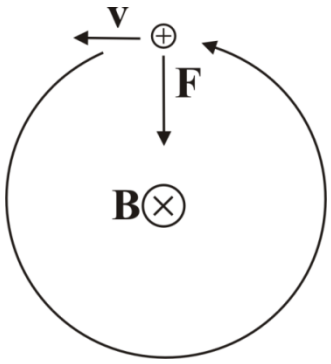


Figure 1.5 Cyclotron motion of a positive ion of charge q moving at velocity **v** in the presence of a constant magnetic field, **B**, which is pointing into the page. The ion moving to the left experiences a downward force, $\mathbf{F} = q(\mathbf{v} \times \mathbf{B})$, $q = ze$, resulting in a counterclockwise 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 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. Ions having a different cyclotron frequency will not be accelerated, and this provides the basis for ion cyclotron resonance mass spectrometry. As the coherently orbiting excited ions passing another opposing pair of electrodes (detection plates), also parallel to the magnetic axis, they induce an alternating current to the plates called image current. 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. Once this transient signal is amplified

and detected, the ions are detected without ever colliding with the electrodes, which makes the detection scheme non-destructive and allows for improved sensitivity and versatility of FT-ICR.

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 1.9). As the cyclotron frequency can be measured with very high precision, the mass accuracy of FT-ICR MS is as high as 1 ppm. The resolving power of FT-ICR 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 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. Therefore, high vacuum (10^{-10} mbar) is necessary in the cell region of FT-ICR MS, to avoid the collision with gas particles and the deactivation of the ions.

1.8 Literature Cited

1. Ghazarian, H.; Idoni, B.; Oppenheimer S. B. *Acta Histochem*. **2011**, *113*, 236-247.
2. Lis, H.; Sharon, N. *Chem*. *Rev*. **1998**, *98*, 637-674.
3. Sharon, N.; Lis, H. *Sci*. *Am*. **1993**, *268*, 82–89.
4. Quiocho, F. A. *Annu*. *Rev*. *Biochem* . **1986**, *55*, 287-315.
5. Holgersson, J.; Gustafsson, A.; Breimer, M. E. *Immunol*. *Cell Biol*. **2005**, *83*, 694-708.

6. Weis, W. I.; Drickamer, K. *Annu*. *Rev*. *Biochem*. **1996**, *65*, 441-473.

7. Sacchettini, J. C.; Baum, L. G.; Brewer, C. F. *Biochemistry* **2001**, *40*, 3009-3015.

8. Utsuno, K.; Uludag, H. *Biophys*. *J*. **2010**, *99*, 201-207.

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

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

11. Herberg, F. W.; Zimmermann, B. **1999** Analysis of protein kinase interactions using biomolecular interaction analysis. In: Protein Phosphorylation - A Practical Approach, D.G. Hardie, ed. (Oxford, UK: Oxford University Press), pp 335-371.

12. Moll, D.; Schweinsberg, S.; Hammann C.; Friedrich W.; Herberg**,** F. W. *Biol*. *Chem*. **2007**, *388*, 163-172.

13. Ng, E. S. M.; Chan, N. W. C., Lewis, D. F.; Hindsgaul, O.; Schriemer, D. C. *Nat*. *Prot*. **2007**, *2*, 1907-1917.

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

15. Horlacher, T.; Seeberger P. H. *Chem*. *Soc*. *Rev*. **2008**, *37*, 1414-1422.

16. Laurent, N.; Voglmeir, J.; Flitsch, S. L. *Chem*. *Comm*. **2008**, *37*, 4400-4412.

17. Liu, Y.; Palma, A. S.; Feizi, T. *Biol*. *Chem*. **2009**, *390*, 647-656.

18. Disney, M. D.; Seeberger, P. H. (2004) *Drug Discov*. *Today*. *TARGETS* **2004**, *3*, 151-158.

19. Ganem, B.; Li, Y. T.; Henion, J. D. *J*. *Am*. *Chem*. *Soc*. **1991**, *113*, 7818-7819.

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

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

22. Jorgensen, T. J. D.; Roepstorff, P.; Heck, A. J. R. *Anal*. *Chem*. **1998**, *70*, 4427-4432.

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

24. Yu, Y. H.; Kirkup, C. E.; Pi, N.; Leary, J. A. *J*. *Am*. *Soc*. *Mass Spectrom*. **2004**, *15*, 1400-1407.

25. Hofstadler, S. A.; Sannes-Lowery, K. A. *Nat*. *Rev*. *Drug Discov*. **2006**, *5*, 585-595.

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

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

28. Sandbhor, M. S.; Soya, N.; Albohy, A.; Zheng, R. B.; Cartmell, J.;Bundle, D. R.; Klassen, J. S.; Cairo, C. W. *Biochemistry* **2011**, *50*, 6753-6762.

29. El-Hawiet, A.; Kitova, E. N.; Kitov, P.; Eugenio, L.; Ng, K. K. S.; Mulvey, G. L.; Dingle, T. C.; Szpacenko, A.; Armstrong, G. D.; Klassen, J. S. *Glycobiology* **2011**, *21*, 1217-1227.

30. Soya, N.; Fang, Y.; Palcic, M. M.; Klassen, J. S. *Glycobiology* **2011**, *21*, 547-552.

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

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

33. Humphries, R. M.; Donnenbberg, M. S.; Strecher, J.; Kitova, E. N.; Klassen, J. S.; Cui, L.; Griener, T. P.; Mulvey, G. L.; Armstrong, G. D. *Mol*. *Microbiol*. **2009**, *72*, 859-868.

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

35. Dingle, T.; Wee, S.; Mulvey, G. L.; Greco, A.; Kitova, E. N.; Sun, J.; Lin, S.; Klassen, J. S.; Palcic, M. M.; Ng, K. K.; Armstrong, G. D. *Glycobiology* **2008**, *18*, 698-706.

36. Kitov, P. I.; Lipinski, T.; Paszkiewicz, E.; Solomon, D.; Sadowska, J. M.; Grant, G. A.; Mulvey, G. L.; Kitova, E. N.; Klassen, J. S.; Armstrong, G. D.; Bundle, D. R. *Angew*. *Chemie Int*. *Ed*. **2008**, *47*, 672-676.

37. 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.

38. Rademacher, C.; Shoemaker, G. K.; Kim, H. S.; Zheng, R. B.; Taha, H.; Liu, C.; Nacario, R. C.; Schriemer, D. C.; Klassen, J. S.; Peters, T.; Lowary, T. L. *J*. *Am*. *Chem*. *Soc*. **2007**, *129*, 10489-10502.

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

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

41. McCammon, M. G.; Scott, D. J.; Keetch, C. A.; Greene, L. H.; Purkey, H. E.; Petrassi, H. M.; Kelly, J. W.; Robinson, C. V. *Structure* **2002**, *10*, 851–863.

42. Yu, Y.; Sweeney, M. D.; Saad, O. M.; Crown, S. E.; Handel, T. M.; Leary, J. A. *J*. *Biol*. *Chem*. **2005**, *280*, 32200-32208.

43. Cummins, L. L.; Chen, S.; Blyn, L. B.; Sannes-Lowery, K. A.; Drader, J. J.; Griffey, R. H.; Hofstadler, S. A. *J*. *Nat*. *Prod*. **2003,** *66*, 1186-1190.

44. Wigger, M.; Eyler, J. R.; Benner, S. A.; Li, W.; Marshall, A. G. *J*. *Am*. *Soc*. *Mass Spectrom*. **2002**, *13*, 1162–1169.

45. Abzalimov, R. R.; Dubin, P. L.; Kaltsahov, I. A. *Anal*. *Chem*. **2007**, *79*, 6055-6063.

46. Cheng, X.; Chen, R.; Bruce, J. E.; Schwartz, B.; Anderson, G. A.; Hofstadler, S. A.; Gale, D. C.; Smith, R. D. *J*. *Am*. *Chem*. *Soc*. **1995**, *117*, 8859-8860.

47. Gao, J.; Cheng, X.; Chen, R.; Sigal, G. B.; Bruce, J. E.; Schwartz, B. L.; Hofstadler, S. A.; Anderson, G. A.; Smith, R. D.; Whitesides, G. M. *J*. *Med*. *Chem*. **1996**, *39*, 1949-1955.

48. Cederkvist, F. H.; Zamfir, A. D.; Bahrke, S.; Eijsink, V. G. H.; Sørlie, M.; Peter-Katalinic, J.; Peter, M. G. *Angew*. *Chem*. *Int*. *Ed*. **2006**, *45*, 2429 –2434.

49. Daneshfar, R.; Kitova, E. N.; Klassen, J. S. *J*. *Am*. *Chem*. *Soc*. **2004** *126*, 4786-4787.

50. Frahm, J. L.; Muddiman, D. C.; Burke, M. J. *J*. *Am*. *Soc*. *Mass Spectrom*. **2005**, *16*, 772-778.

51. Benesch, J. L. P.; Sobott, F.; Robinson, C. V. *Anal*. *Chem*. **2003**, *75*, 2208–2214.

52. Wang, G.; Abzalimov, R. R.; Kaltashov, I. A. *Anal*. *Chem*. **2011**, *83*, 2870-2876.

53. Ganem, B.; Henion, J. D. *Bioorg*. *Med*. *Chem*. **2003**, *11*, 311-314.

54. Krishnaswamy, S. R.; Williams, E. R.; Kirsch, J. F. *Protein Sci*. **2006,** *15*, 1465-1475.

55. Peschke, M.; Verkerk, U. H.; Kebarle, P. *J*. *Am*. *Soc*. *Mass Spectrom*. **2004,** *15*, 1424-1434.

56. Gross, J., Xavier, A., Gross, M.: RNA Neomycin Complexes Analyzed by Electrospray and UV-MALDI-MS. Abstracts of Papers, 49th ASMS Conference on Mass Spectrometry and Allied Topics; Chicago, IL; June **2001**.

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

58. Wilcox, J. M.; Rempel, D. L.; Gross, M. L. *Anal. Chem.* **2008**, *80*, 2365–2371.

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

60. Gabelica, V.; Rosu, F.; De Pauw, E. *Anal. Chem.* **2009**, *81*, 6708-6715.

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

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

63. Sun, J.; Kitova, E. N.; Klassen, J. S. *Anal. Chem.* **2007**, *79*, 416-425.

64. Xie, Y.; Zhang, J.; Yin, S.; Loo, J. A. *J. Am. Chem. Soc.* **2006**, *128*, 14432–14433.

65. Bagal, D.; Kitova, E. N.; Liu, L.; El-Hawiet, A.; Schnier, P. D.; Klassen, J. S. *Anal. Chem.* **2009**, *81*, 7801-7806.

66. Clark, S. M.; Konermann, L. *Anal. Chem.* **2004**, *76*, 7077-7083.

67. Robinson, C. V.; Chung, E. W.; Kragelund, B. B.; Knudsen, J.; Aplin, R. T.; Poulsen, F. M.; Dobson, C. M. *J. Am. Chem. Soc.* **1996**, *118*, 8646-8653.

68. Hossain, B. M.; Konermann, L. *Anal. Chem.* **2006**, *78*, 1613-1619.

69. Hunter, E. P.; Lias, S. G. *Phys. Chem. Ref. Data* **1998**, *27*, 413-656.

70. Taft, R. W.; Anvia, F.; Taagepera, M.; Catalan, J.; Elguero, J. *J. Am. Chem. Soc.* *1***986**, *108*, 3237-3239.

71. Kitova, E. N.; Seo, M.; Roy, P. N.; Klassen, J. S. *J. Am. Chem. Soc.* **2008**, *130*, 1214-1216.

72. Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C., Ferguson, L. D.; Alice, M. B. *J*. *Chem*. *Phys*. **1968,** *49*, 2240–2249.

73. Wang, W.; Kitova, E. N.; Klassen, J. S. *Anal*. *Chem*. **2005,** *77*, 3060-3071.

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

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

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

77. Daubenfeld, T.; Bouin, A. P.; van der Rest, G. *J*. *Am*. *Soc*. *Mass Spectrom*. **2006**, *17*, 1239-1248.

78. Shimon, L.; Sharon, M.; Horovitz, A. *Biophys*. *J*. **2010,** *99*, 1645-1649.

79. Lane, L. A.; Ruotolo, B. T.; Robinson, C. V.; Favrin, G.; Benesch, J. L. P. *J*. *Mass Spectrom*. **2009**, *283*, 169-177.

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

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

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

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

84. Hernandez, H.; Robinson, C. V. *Nature Prot*. **2007**, *2*, 715–726.

85. Iavarone, A. T.; Udekwu, O. A.; Williams, E. R. *Anal*. *Chem*. **2004**, *76*, 3944–3950.

86. Yang, P.; Cooks, R. G.; Ouyang, Z.; Hawkridge, A. M.; Muddiman, D. C. *Anal*. *Chem*. **2005,** *77*, 6174–6183.

87. Ferguson, C. N.; Benchaar, S. A.; Miao, Z.; Loo, J. A.; Chen, H. *Anal*. *Chem*. **2011**, *83*, 6468-6473.

88. Wortmann, A.; Jecklin, M. C.; Touboul, D.; Badertscher, M.; Zenobi, R. *J*. *Mass Spectrom*. **2008**, *43*, 600-608.

89. Jecklin, M. C.; Touboul, D.; Jain, R.; Toole, E. N.; Tallarico, J.; Drueckes, P.; Ramage, P.; Zenobi, R. *Anal*. *Chem*. **2009**, *81*, 408–419.

90. Tang, L.; Kebarle, P. *Anal*. *Chem*. **1993**, *65*, 972A-986A.

91. Kebarle1, P.; Verkerk, U. H. *Mass Spectrom*. *Rev*. **2009**, *28*, 898– 917.

92. Taylor, G. I. *J Fluid Mech* **1965**, *2*, 1-15.

93. (a) Iribarne, J. V.; Thomson, B. A. *J*. *Chem*. *Phys*. **1976**, *64*, 2287-2294; (b) Iribarne, J. V.; Thomson, B. A. *J*. *Chem*. *Phys*. **1979**, *71*, 4451-4463.

94. Hogan, C. J.; Carroll, J. A.; Rohrs, H. W.; Biswas, P., Gross, M. L. *J Am Chem Soc*. **2008**, *130*,6926–6927.

95. De Hoffmann, E.; Stroobant, V. Mass Spectrometry Principles and Applications 3th Ed. 2007, New York: John Wiley & Sons.

96. Wollnik, H. *Mass Spectrom*.*Rev*. **1993**, *12*, 89-114.

97. Weickhardt, C.; Moritz, F.; Grotemeyer, *J*. *Mass Spectrom*.*Rev*. **1996**, *15*, 139-162.

98. Mamyrin, B. A., *Int*. *J*. *Mass spectrom*. **2001**, *206*, 251-266 .

99. El-Aneed, A.; Cohen, A.; Banoub, *J*. *Appl Spectrosc Rev* **2009**, *44*: 210–230.

100. Creaser, C. S. ; Griffiths, J. R.; Bramwell, C. J.; Noreen, S.; Hill, C. A. ; Paul Thomas, C. L. *Analyst* **2004**, *129*, 984 – 994.

101. Karasek, F. W. *Anal*. *Chem*. **1974**, *46*, 710A.

102. McDaniel, E. W.; Mason, E. A. The Mobility and Diffusion of Ions in Gases, John Wiley & Sons, New York, USA, 1973.

103. Kanu, A. B.; Dwivedi, P.; Tam, M.; Matz, L.; Hill, H. H. *J. Mass Spectrom.* **2008**, *43*, 1-22.

104. Pringle, S. D.; Giles, K.; Wildgoose, J. L.; Williams, J. P.; Slade, S. E.; Thalassinos, K.; Bateman, R. H.; Bowers, M. T.; Scrivens, J. H. *Int. J. Mass spectrom.* **2007**, *261*, 1-12.

105. Sergio, A..; Manel, A..; Marcelo, B. *Anal. Chim. Acta* **2011**, *703*, 114– 123.

106. Cohen, M. J.; Karasek, F. W. *J. Chromatogr. Sci.* **1970**, *8*, 330-339.

107. Sacristan, E.; Solis, A. A. *IEEE Trans. Instrum. Meas.* **1998**, *47*, 769-777.

108. Buryakov, I. A.; Krylov, E. V.; Nazarov, E. G.; Rasulev, U. K. *Int. J. Mass Spectrom. Ion Processes* **1993**, *128*, 143-152.

109. Giles, K.; Pringle, S. D.; Worthington, K. R.; Little, D.; Wildgoose, J. L.; Bateman, R. H. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2401-2406.

110. Amster, I. J. *J. Am. Soc. Mass Spectrom.* **1996**, *31*, 1325-1337.

111. Buchanan, M. V.; Hettich, R. L. *Anal. Chem.* **1993**, *65*, A245-A259.

112. Marshall, A. G.; Grosshans, P. B. *Anal. Chem.* **1991**, *63*, A215-A229.

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

Chapter 2

Binding of *Clostridium difficile* Toxins to Human Milk Oligosaccharides[*+‡]

2.1 Introduction

Clostridium difficile is a Gram-positive, spore forming, strict anaerobic bacterium responsible for a variety of toxin-mediated gastrointestinal diseases that range in severity from antibiotic-associated diarrhea to pseudomembranous colitis.[1,2] The emergence of a new and more virulent strain in North America and Europe has been linked to increased morbidity and mortality. Although *C*. *difficile* can produce up to six different toxins, the main virulence factors are the two exotoxins, toxin A (TcdA) and toxin B (TcdB).[3,4] These two toxins, which share 47% amino acid sequence identity, belong to the large clostridial glucosylating toxin family.[4] Both toxins catalyze the transfer of glucose onto the Rho family of GTPases, leading to a disruption of the cytoskeleton and cell death.[5] Although both TcdA and TcdB share a common glucosyltransferase activity and overall structure, differences in structure, substrate specificity and receptor binding appear to contribute towards different cytotoxic mechanisms.[4-8] The two toxins appear to exert complementary effects to synergistically disrupt the intestinal epithelium during pathogenesis.

---

[*] A version of this chapter has been published: El-Hawiet, A.; Kitova, E. N.; Kitov, P.; Eugenio, L.; Ng, K. K. S.; Mulvey, G. L.; Dingle, T. C.; Szpacenko, A.; Armstrong, G. D.; Klassen, J. S. *Glycobiology* **2011**, *21*, 1217-1227.

[+] Protein expression and purification were done by Eugenio, L.(University of Calgary)

[‡] Vero cytotoxicity neutralization assays were done by Mulvey, G. L.and Dingle, T. C (University of Calgary)

Like all members of the clostridial toxin group, TcdA and TcdB are large (308 and 250 kDa, respectively) single-subunit polypeptides, whose structures appear to be organized into three regions: (i) an N-terminal glucosyltransferase domain; (ii) a central region containing cysteine and aspartyl proteolytic activities, as well as a hydrophobic region, which is important for translocating the toxins across the cell membrane; and (iii) a highly repetitive C-terminal region, which appears to be primarily responsible for receptor binding.[8-12] Although *C. difficile* infections can usually be controlled by treatment with broad-spectrum antibiotics, like metronidazole and vancomycin, existing therapeutic approaches are not effective for treating novel hypervirulent and drug-resistant strains, as well as many cases of relapse or reinfection due to the continued disruption of normal bacterial flora following antibiotic treatment.[13] As a result, alternative therapeutic strategies are required to prevent *C. difficile* from colonizing the intestinal tract and to neutralize the cytotoxic effects of TcdA and TcdB. It has been proposed that such a therapy may involve host cell receptor analogs in various forms that are able to competitively inhibit TcdA and TcdB from binding to the surface of human intestinal epithelial cells.[14] The rationale behind this approach is to provide toxins with decoy ligands in the gastrointestinal tract that will divert them from their native receptors on the host cell surface, thus sequestering the toxins and facilitating their elimination from the body.

The specific functional receptors for TcdA and TcdB toxins in humans have yet to be positively identified.[15-17] Currently, the only known native receptor for TcdA is the trisaccharide α-D-Gal(1,3)β-D-Gal(1,4)β-D-GlcNAc, which is found on the surface of rabbit erythrocytes, hamster brush border membranes, bovine thyroglobulin

and both Immunoglobulin (Ig) and non-Ig components of human milk.[15-17] Recently, it was shown that the related trisaccharide αGal(1,3)βGal(1,4)βGlc and its analogs bind specifically, albeit weakly, to fragments of TcdA and TcdB.[18] TcdA also binds to Lewis X, Y, and I glycan sequences, which are expressed on the surface of human intestinal epithelial cells.[19,20]

Human milk oligosaccharides (HMOs) are known to protect newborns from a variety of infectious diseases.[21] For example, HMOs have been shown to inhibit the attachment of *Streptococcus pneumoniae*, *Norwalk-like* virus and *Haemophilus influenzae* to host cells.[22,23] Furthermore, fucosylated oligosaccharides from human milk have been found to protect infants from the heat stable toxin of *E. coli* and to prevent the binding of *Campylobacter jejuni* to its receptor in human epithelial cells.[24,25] Additionally, the trisaccharide 3′-sialyllactose inhibits the binding of *Helicobacter pylori* to the gastrointestinal epithelium and shows protection against cholera toxin-induced diarrhea.[26-28] While protection may be due to the prebiotic characteristics of HMOs, it is believed to result primarily from inhibition of binding of pathogens to host cells due to the similarity of HMOs to epithelial cell surface carbohydrates.

In the present study, the potential of HMOs as inhibitors of *C. difficile* TcdA and TcdB was explored. The binding of twenty-one HMOs, which represent the most abundant oligosaccharides in human milk, to fragments of TcdA and TcdB was investigated using the direct electrospray ionization mass spectrometry (ESI-MS) assay. Molecular docking simulations were carried out to elucidate the molecular basis of HMO recognition by the toxin fragments. Cytotoxicity neutralization assays were also performed to investigate the inhibitory potential of HMOs on TcdA and TcdB.

2.2 Experimental

2.2.1 Proteins

The TcdA-A2 subfragment (A2, MW 29590 Da) and TcdB-B1 subfragment (B1, MW 14860 Da) were expressed in *Escherichia coli* and purified as described previously.[29,30] Purified samples of TcdA-A2 and TcdB-B1 solutions were stored at -80 ^{o}C. Lysozyme (Lyz, MW 14310 Da) and α-lactalbumin (LA, MW 14200 Da), which served as reference proteins (P_{ref}) for the ESI-MS binding assays, were purchased from Sigma–Aldrich Canada (Oakville, ON) and used without further purification.

2.2.2 Carbohydrates

The HMOs library consisted of lactose (β-D-Gal(1-4)β-D-Glc), MW 342 Da, (**L1**); 2′-fucosyl-lactose, α-D-Fuc(1-2)β-D-Gal(1-4)β-D-Glc, MW 488 Da, 2FL (**L2**); 3-fucosyl-lactose, β-D-Gal(1-4)[α-D-Fuc(1-3)]β-D-Glc, MW 488 Da, 3FL (**L3**); difucosyl-lactose, α-D-Fuc(1-2)β-D-Gal(1-4)[α-D-Fuc(1-3)]β-D-Glc, MW 634 Da, LDFT (**L4**); 3′-sialyl-lactose, 5-Acetyl-α-Neu(2-3)β-D-Gal(1-4)β-D-Glc, MW 633 Da, 3SL (**L5**); 6′-sialyl-lactose, 5-Acetyl-α-Neu(2-6)β-D-Gal(1-4)β-D-Glc, MW 633 Da, 6SL (**L6**); 3′-sialyl-3-fucosyl-lactose, 5-Acetyl-α-Neu(2-3)β-D-Gal(1-4)[α-D-Fuc(1-3)]β-D-Glc, MW 779 Da, 3′NeuAc-3FL (**L7**); lacto-N-tetraose, β-D-Gal(1-3)β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc, MW 708 Da, LNT (**L8**); lacto-N-fucopentaose I, α-D-Fuc(1-2)β-D-Gal(1-3)β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc, MW 853 Da, LNFI (**L9**); lacto-N-fucopentaose II, β-D-Gal(1-3)[α-D-Fuc(1-4)] β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc, MW 853 Da, LNFII (**L10**); lacto-N-difucohexaose I, α-D-Fuc(1-2)β-D-Gal(β1-3)[α-D-Fuc(1-4)] β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc, MW 1000

Da, LNDI (**L11**); difucosyl-lacto-N-hexaose(a), β-D-Gal(1-4)[α-D-Fuc(1-3)]β-D-GlcNAc(1-6)[α-D-Fuc(1-2) β-D-(1-3) β-D-GlcNAc(1-3)] β-D-Gal(1-4)β-D-Glc, MW 1365 Da, F2-LNH a (**L12**); sialyl-lacto-N-tetraose a, 5-Acetyl-α-Neu(2-3)β-D-Gal(1-3)β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc, MW 998 Da, LST a (**L13**); sialyl-lacto-N-tetraose b, 5-Acetyl-α-Neu(2-6)[β-D-Gal(1-3)] β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc, MW 998 Da, LST b (**L14**); disialyl-lacto-N-tetraose, 5-Acetyl-α-Neu(2-3)β-D-Gal(1-3)[5-Acetyl-α-Neu(2-6)]β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc, MW 1290 Da, disialyl-LNT (**L15**); sialyl-fucosyl-lacto-N-tetraose, 5-Acetyl-α-Neu(2-3)β-D-Gal(1-3)[α-D-Fuc(1-4)]β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc, MW 1145 Da, sialyl-Le^{a} or Sia-LNF III (**L16**); sialyl-lacto N-fucopentaose V, α-D-Fuc(1-2)β-D-Gal(1-3)[5-Acetyl-α-Neu(2-6)]β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc, MW 1145 Da, Sia-LNF V (**L17**); lacto-N-neo-tetraose, β-D-Gal(1-4)β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc, MW 708 Da, LnNT (**L18**); lacto-N-fucopentaose III, β-D-Gal(1-4)[α-D-Fuc(1-3)]β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc, MW 853 Da, LNFIII (**L19**); difucosylpara-lacto-N-hexaose,β-D-Gal(1-3)[Fuc(α1-4)]β-D-GlcNAc(1-3)β-D-Gal(1-4)[α-D-Fuc(1-3)]β-D-GlcNAc(1-3)]β-D-Gal(1-4)β-D-Glc, MW 1365 Da, Lea/Lex (**L20**); sialyl-lacto-N-tetraose c, 5-Acetyl-α-Neu(2-6)β-D-Gal(1-4)β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc, MW 998 Da, LST c (**L21**). The HMOs were purchased from Sigma–Aldrich (Oakville, ON) (**L1**) and IsoSep AB (Sweden) (**L2-L21**). Stock solutions were prepared by dissolving all ligands with ultrafiltered water (Milli-Q, Millipore) at a concentration of 1 mM and stored at -20 ^{o}C until used.

2.2.3 Isolation of HMOs from human milk

Human milk donations were obtained from the Alberta Children's Hospital, Calgary, Alberta, Canada. Oligosaccharides were extracted as previously described.[31] Raw milk (1 L) was centrifuged at 5,000 ×g for 30 min at 4°C, and the fat was removed. Ethanol (2 L) was added, and the solution was incubated overnight at 24°C. The precipitate was removed by centrifugation at 5,000 ×g for 30 min at 4°C, and the solvent was removed by rotary evaporation. The carbohydrate fraction was dissolved in 5 mL of water and the solution was passed through a Bio Gel P-2 (Extra fine, <45 μm; Bio Rad Laboratories, Hercules, CA, USA) column (2.6 X 100 cm). Elution was performed with 100 mM ammonium acetate at a flow rate of 26 mL/h, and the elution profile was recorded with a refractive Index detector (Waters, differential refractometer R401). A total of 6 (I-VI) HMO fractions (7.5 mL volume each) were collected and freeze-dried. Representative ESI mass spectra acquired for aqueous solutions of 1mg/100μL of each HMO fractions are given in Figure 2.1. Mass and composition of the oligosaccharides detected in the isolated HMO fractions are listed in Table 2.1.

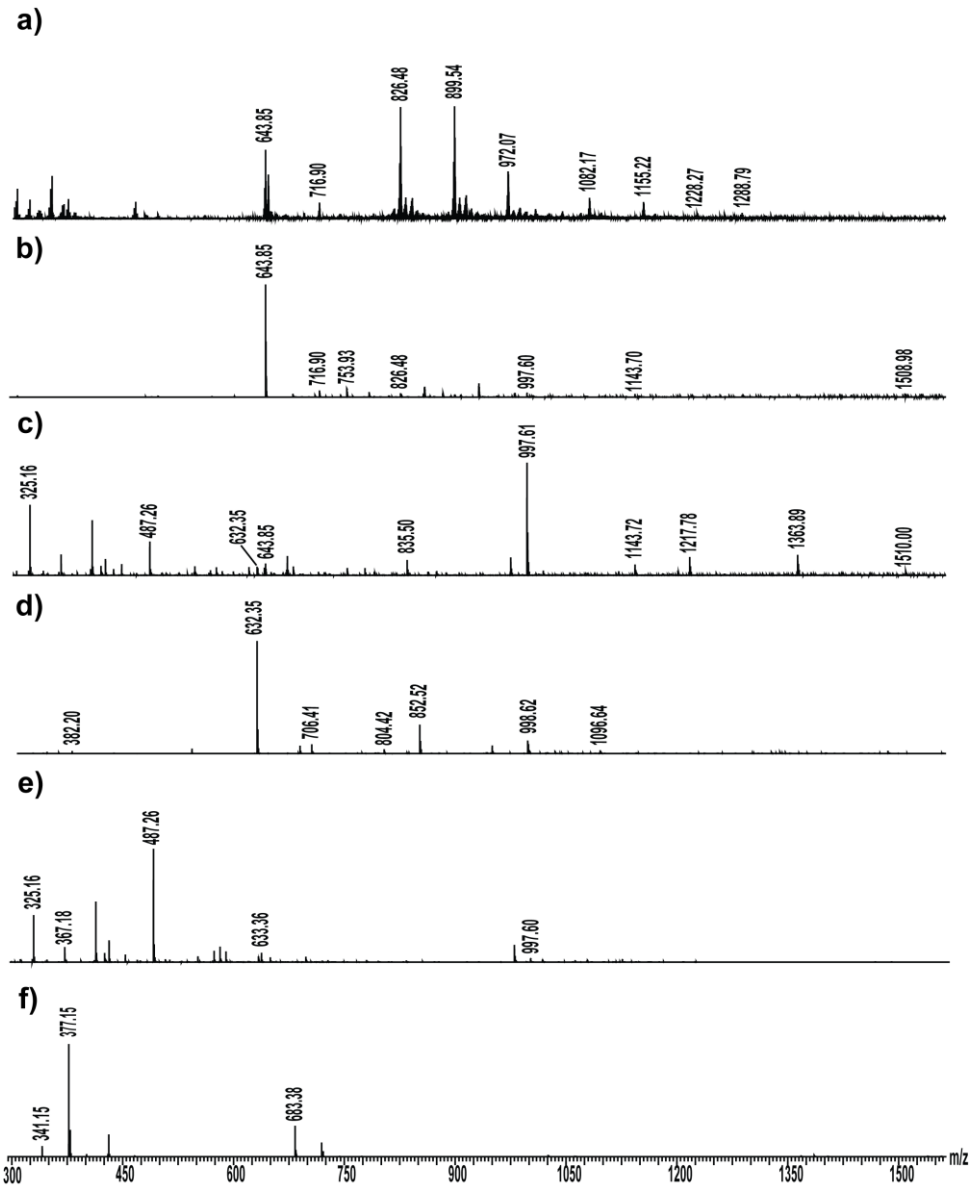


Figure 2.1: ESI mass spectra obtained in negative ion mode with a Synapt HDMS G2 Q-TOF (Waters, Manchester, UK) for (a) HMO fraction I, (b) fraction II, (c) fraction III, (d) fraction IV, (e) fraction V, (f) fraction VI obtained from human milk using size exclusion chromatography separation.

Table 2.1: Mass and composition of the oligosaccharides detected by ESI-MS from the HMO fractions

HMO Fraction	m/z	Mass (measured)	Mass (theoretical)	Monomer composition
Fraction I	643.85	1289.75	1289.42	3Hex:1HexNAc:2NeuAc
	716.90	1435.84	1436.53	5Hex:3HexNAc
	826.48	1654.96	1654.43	4Hex:2HexNAc:2Fuc:1NeuAc, 4Hex:2HexNAc:2NeuAc, 4Hex:2HexNAc:4Fuc
	899.54	1801.08	1800.83	4Hex:2HexNAc:1Fuc:2NeuAc, 6Hex:4HexNAc
	972.07	1946.18	1946.91	4Hex:2HexNAc:3NeuAc 4Hex:2HexNAc:2Fuc:2NeuAc
	1082.17	2166.34	2166.62	5Hex:3HexNAc:1Fuc:2NeuAc, 5Hex:3HexNAc:3Fuc:1NeuAc, 7Hex:5HexNAc
	1155.22	2313.0	2312.41	5Hex:3HexNAc:3NeuAc, 5Hex:3HexNAc:2Fuc:2NeuAc, 5Hex:3HexNAc:4Fuc:1NeuAc, 7Hex:5HexNAc:1Fuc
	1228.27	2458.54	2458.72	5Hex:3HexNAc:1Fuc:3NeuAc, 5Hex:3HexNAc:3Fuc:2NeuAc, 5Hex:3HexNAc:5Fuc:1NeuAc, 7Hex:5HexNAc:1NeuAc 7Hex:5HexNAc:2Fuc
	1288.79	1289.79	1289.42	3Hex:1HexNAc:2NeuAc
Fraction II	643.85	1289.83	1289.42	3Hex:1HexNAc:2NeuAc
	716.90	1435.80	1436.53	5Hex:3HexNAc
	753.93	1509.88	1510.63	4Hex:2HexNAc:1Fuc:1NeuAc**,** 4Hex:2HexNAc:3Fuc
	826.48	1654.96	1654.43	4Hex:2HexNAc:2Fuc:1NeuAc, 4Hex:2HexNAc:2NeuAc, 4Hex:2HexNAc:4Fuc
	997.60	998.61	998.34	3Hex:1HexNAc:1NeuAc, 3Hex:1HexNAc:2Fuc
	1143.70	1144.70	1145.42	3Hex:1HexNAc:1Fuc:1NeuAc
	1508.98	1509.98	1510.63	4Hex:2HexNAc:1Fuc:1NeuAc**,**

				4Hex:2HexNAc:3Fuc
Fraction III	325.16	325.16	326.77	1Hex:1Fuc
	487.26	488.26	488.17	2Hex:1Fuc
	632.35	633.35	633.21	2Hex:2Fuc, 2Hex: 1NeuAc
	643.85	1289.74	1289.42	3Hex:1HexNAc:2NeuAc
	835.50	836.50	837.31	2Hex:1HexNAc:1NeuAc
	997.61	998.61	998.34	3Hex:1HexNAc:1NeuAc, 3Hex:1HexNAc:2Fuc
	1143.72	1144.72	1145.42	3Hex:1HexNAc:1Fuc:1NeuAc
	1217.78	1218.78	1219.54	4Hex:2HexNAc:1Fuc
	1363.89	1364.89	1365.53	4Hex:2HexNAc:2Fuc, 4Hex:2HexNAc:1NeuAc
	1510.00	1511.00	1511.63	4Hex:2HexNAc:1Fuc:1NeuAc**,** 4Hex:2HexNAc:3Fuc
Fraction IV	382.2	383.20	382.12	1Hex:1HexNAc
	632.35	633.35	633.21	2Hex:2Fuc, 2Hex:1NeuAc
	706.41	707.41	707.25	3Hex:1HexNAc
	804.42	782.42	782.32	2Hex:1Fuc:1NeuAc
	852.52	853.52	853.33	3Hex:1HexNAc:1Fuc
	998.62	999.63	998.34	3Hex:1HexNAc:1NeuAc, 3Hex:1HexNAc:2Fuc
	1096.64	1074.64	1074.41	4Hex:2HexNAc
Fraction V	325.16	326.16	326.11	1Hex:1Fuc
	367.18	345.18	344.62	2Hex
	487.26	488.26	488.17	2Hex:1Fuc
	633.36	634.36	633.81	2Hex:2Fuc, 2Hex: 1NeuAc
	997.60	998.60	998.34	3Hex:1HexNAc:1NeuAc, 3Hex:1HexNAc:2Fuc
Fraction VI	341.15	342.15	342.12	2Hex
	377.15	341.15	342.45	2Hex
	683.38	684.38	684.22	2Hex dimer

2.2.4 Direct ESI-MS assay

Apparent association constants ($K_{a,app}$) for the fragments TcdA-A2 and TcdB-B1 binding to the library of twenty one HMOs (**L1** – **L21**) were evaluated using the ESI-MS assay. Complete details of the experimental methodology and data analysis are described elsewhere[32,33] and only a brief overview is given here. The ESI-MS measurements were carried out using a 9.4T Apex II Fourier-transform ion cyclotron resonance (FTICR) MS (Bruker-Daltonics, Billerica, MA). Prior to analysis, the TcdA-A2 and TcdB-B1 solutions were diluted with 50 mM ammonium acetate (pH 7.2) and concentrated using Amicon Ultra-4 centrifugal filters with a molecular weight cut-off of 10,000 Da (Millipore). The concentrations of the TcdA-A2 and TcdB-B1 solutions were measured by UV absorption. Each ESI solution was prepared from stock solutions of protein (TcdA-A2 and TcdB-B1) and one of carbohydrate ligands (**L1** – **L21**). Lysozyme (Lyz) and α-lactalbumin (La) were used as reference proteins (P_{ref}) to distinguish specific from nonspecific ligand binding with TcdA-A2 and TcdB-B1, respectively, during the ESI-MS measurements.

2.2.5 Docking simulations

Automated molecular docking simulations were conducted with AutoDock Vina 1.1.1[34] in conjunction with the MGL Tools 1.5.4 graphical interface (Scripps Research Institute, La Jolla, CA). The crystal structure of the complex between TcdA-A2 and α-D-Gal(1,3)β-D-Gal(1,4)β-D-GlcNAc[29] (Protein Databank entry 2G7C) was used without modifications except for the addition of polar hydrogen atoms (hydrogens bound to hetero atoms) as required by AutoDock. The original bound ligand and all water molecules were removed, followed by addition of polar hydrogens to the crystal structure using AutoDock Tools. For TcdB-B1, a homology model was generated by

Modeller[35] with template-based refinement of the binding site as described below. Ligand structures were built using the program Insight II 2005 (Accelrys Inc.) and their energies were minimized using the standard AMBER force field with Homans' parameters for carbohydrates.[36] A distance-dependent dielectric constant (= 4r) was used and the 1 – 4 nonbonded interactions were scaled by 0.5. The Grid box was centered on the ring oxygen of the β-galactose moiety of the original oligosaccharide ligand found in the crystal structure while box parameters were set at 30 Å in each dimension. The proteins were regarded as rigid, while all non-ring bonds in ligands were set as active (flexible). Alternatively, some or all Φ dihedral angles of inter-glycosidic anomeric bonds were set as inactive (rigid) in order to facilitate finding binding modes in agreement with exo-anomeric effect. The energy range for the docked poses was set at 2.75 kcal mol^{-1} and the number of poses was set at 20. All other docking parameters were set to their default values. When a single docking experiment did not result in a pose in which all the Φ dihedral angles were consistent with the exo-anomeric effect, docking were repeated several times with different random seeds. These poses, in which all Φ dihedral angles were consistent with exo-anomeric effect were retained, were then further refined by performing geometry optimization (with the protein structure fixed) using the AMBER force field.

2.2.6 Vero cytotoxicity neutralization assays

Each HMO fraction (I-VI) was 3 fold serial diluted in phosphate-buffered (pH 7.2) physiological saline (PBS). The HMO dilutions were then admixed with purified TcdA or TcdB holotoxins[14] diluted in PBS to their CD_{100} concentration; the minimum concentration resulting in a 100% cytopathic effect in the Verocytotoxicity assay. From each of these samples, 20 μL was transferred to a 96-well microtiter plate containing confluent Vero cell monolayers cultivated in MEM tissue culture growth medium excluding fetal bovine serum (FBS). The final concentration of the HMO fractions in the first well of the Vero cell plate was 2.72 mg/ml. Wells containing TcdA or TcdB specific polyclonal rabbit antisera serial diluted in PBS admixed with TcdA or TcdB served as positive inhibition controls and wells containing PBS alone served as the negative inhibition controls. The microtiter plates were then incubated for 4 h at 37°C before the medium in each well was removed and replaced with fresh MEM supplemented with 10% FBS. The plates were incubated for an additional 48 h and cell viability was subsequently assessed by the conventional Giemsa staining technique and the results were recorded using a microtitre plate reader set to an absorbance of 630 nm.

2.3 Results and discussion

2.3.1 ESI-MS binding measurements

The direct ESI-MS assay was used to test for specific binding between the A2 fragment of TcdA and the B1 fragment of TcdB and each of the twenty-one HMOs (**L1**-**L21**) and to quantify their affinities at pH 7 and 25 °C. A detailed description of the ESI-MS results obtained for **L8**, which binds to both toxin fragments and served as

a model ligand for establishing appropriate experimental and instrumental conditions for the binding measurements, is given below followed by a summary of the results obtained for the other HMOs.

Shown in Figures 2.2a and 2.2b are representative ESI mass spectra acquired for solutions of aqueous ammonium acetate (10 mM), A2 (75 μM), Lyz (12 μM) and **L8** at 50 μM and 100 μM, respectively. As noted above, Lyz served as P_{ref} for the binding measurements performed on A2. Inspection of the ESI mass spectra reveals signals corresponding to protonated ions of free (unbound) A2 fragment, as well as the A2 bound to one or two molecules of **L8**, *i.e.*, $(A2 + q\mathbf{L8})^{n+}$ where $q = 0 – 2$ and n = 9 - 12. Ions corresponding to unbound and bound P_{ref} ions were also detected, *i.e.*, $(P_{ref} + q\mathbf{L8})^{n+}$ where $q = 0$ - 2 and n = 7 - 9, indicating that nonspecific binding of **L8** to A2 occurred during the ESI process and contributed to the mass spectrum. The distributions of **L8** molecules bound to A2 and to Lyz determined from the mass spectra (Figures 2.2a and 2.2b) are shown in Figures 2.2c and 2.2d. Also shown are the distributions of **L8** bound to A2 following correction for nonspecific binding. It can be seen that, under these solution conditions, A2 binds a maximum of one molecule of **L8**. Notably, control experiments, which involved varying the ion source conditions, confirmed that the measured distributions of bound **L8** (after correction for nonspecific binding) were not influenced by in-source (gas phase) dissociation.[33] Measurements were also performed using higher concentrations of **L8**, up to 200 μM. However, no ions corresponding to specific (A2 + 2(**L8**)) complex were detected. The average $K_{a,app}$ obtained from these measurements, after correction for nonspecific binding, is 1500 ± 500 M^{-1}.

The absence of signal corresponding to the (A2 + 2(**L8**)) complex appears, at first glance, to be at odds with the x-ray crystal structure of A2 bound to a synthetic derivative of the natural carbohydrate receptor, which demonstrated that the A2 fragment has two equivalent binding sites for the α-D-Gal(1,3)β-D-Gal(1,4)β-D-GlcNAc trisaccharide.[29] However, based on the measured $K_{a,app}$ and an initial A2 concentration of 75 μM, a **L8** concentration in excess of 300 μM would be required to produce a detectable concentration of the (A2 + 2(**L8**)) complex. It was found that **L8** concentrations >200 μM led to significant protein signal suppression and extensive nonspecific ligand binding. As a result, the ESI-MS binding measurements were restricted to **L8** concentrations ≤ 200 μM.

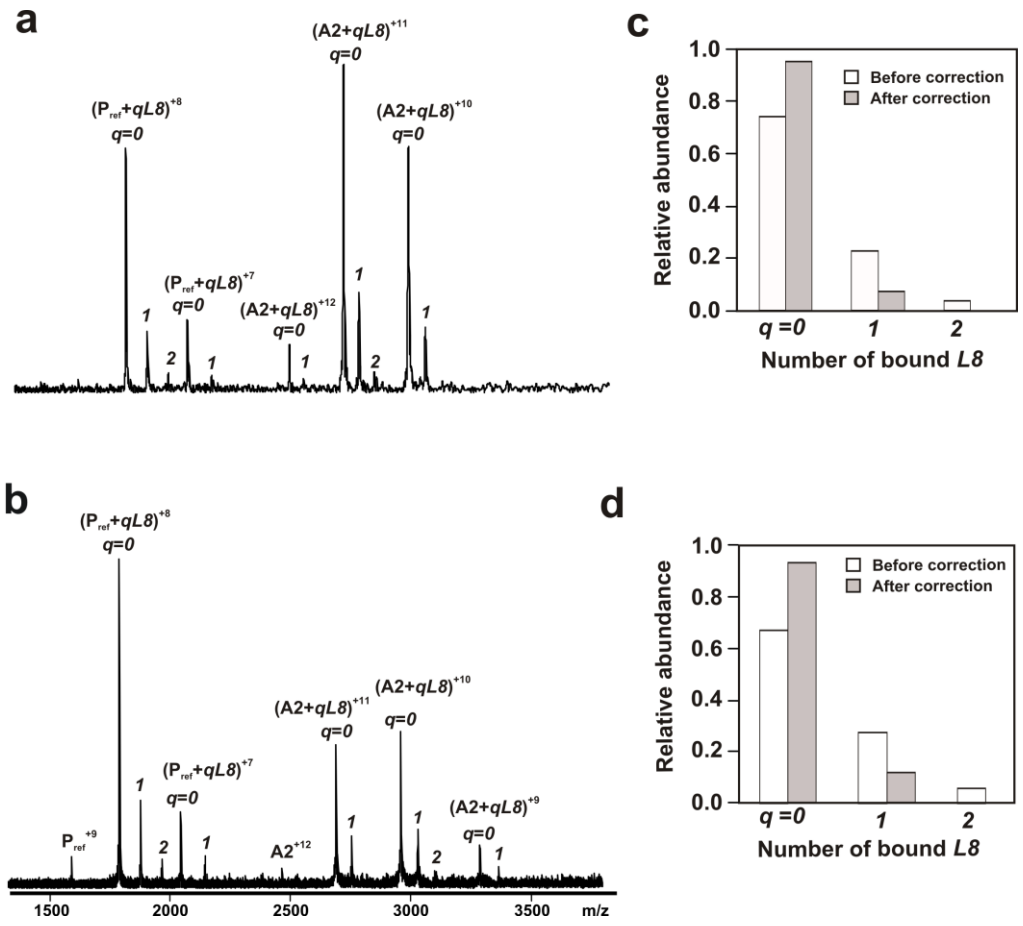


Figure 2.2 ESI mass spectra of aqueous solutions containing (a) 75 μM A2 and 50 μM **L8**, (b) 75 μM A2 and 100 μM **L8**, at pH 7 and 25°C. A P_{ref} (12 μM) was added to each solution to quantify the extent of nonspecific protein-ligand binding during the ESI process. (c) and (d) Normalized distributions of **L8** bound to proteins determined from ESI mass spectra acquired for the solutions described in (a) and (b), respectively.

Shown in Figures 2.3a and 2.3b are ESI mass spectra acquired for aqueous solutions of ammonium acetate (10 μM), B1 (20 μM), LA (8 μM) and **L8** at 50 μM and 150 μM, respectively. It should be noted that LA was used as P_{ref} for these measurements because the addition of Lyz (the P_{ref} used with A2) causes precipitation of B1. Protonated ions corresponding to the free and ligand-bound B1 were detected, *i.e.*, $(B1 + q\mathbf{L8})^{n+}$ where q = 0 - 2 and n = 5 and 6. Free and bound P_{ref} ions were also detected, *i.e.*, $(P_{ref} + q\mathbf{L8})^{n+}$ where $q = 0 - 2$ and n = 6 and 7, indicating that nonspecific binding of **L8** to B1 during the ESI process contributed to the mass spectrum. Shown in Figures 2.3c and 2.3d are the distributions of **L8** bound to B1 obtained from the mass spectra shown in Figure 2.3a and 2.3b, respectively, before and after correction for nonspecific binding. According to these results, and those obtained at different ligand concentrations, the B1 fragment binds a single molecule of **L8**, which is consistent with the presence of a single carbohydrate binding site. The average $K_{a,app}$ value is $1000 \pm 500\ M^{-1}$.

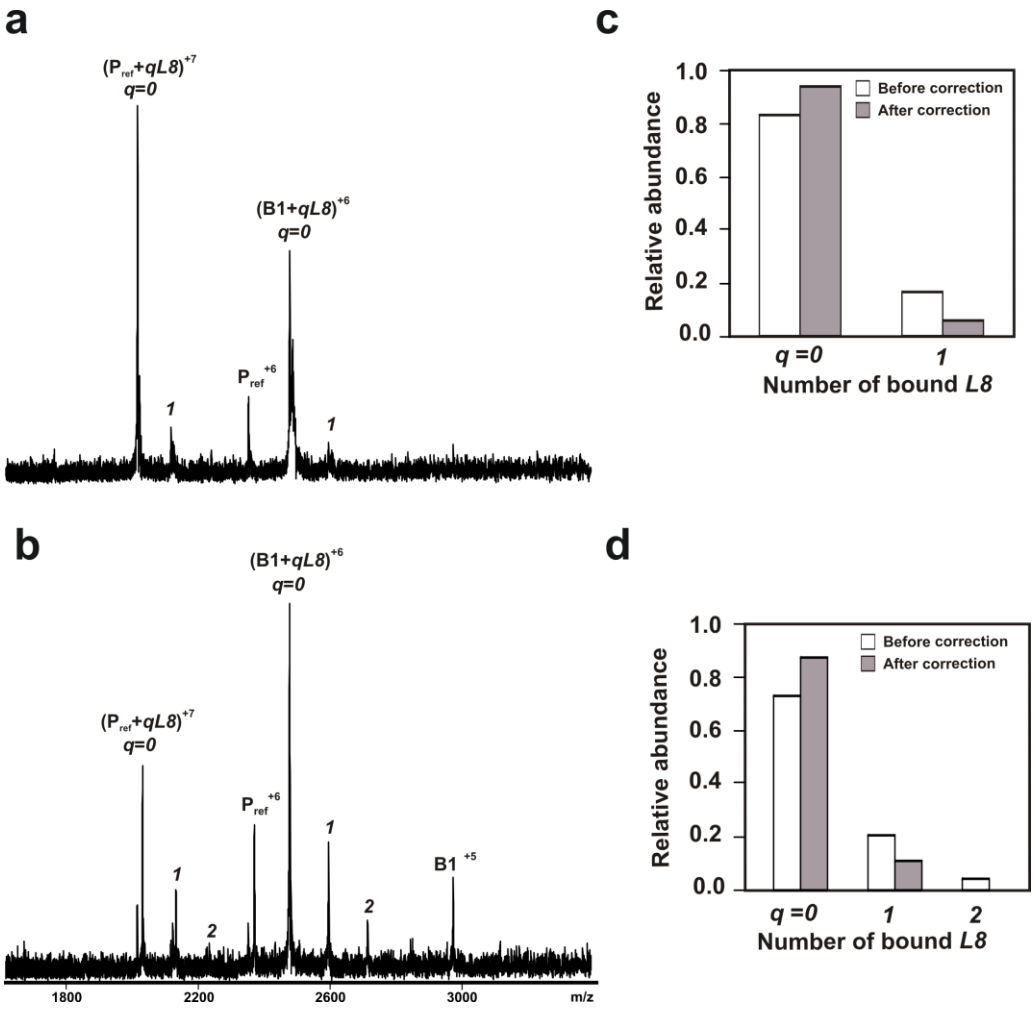
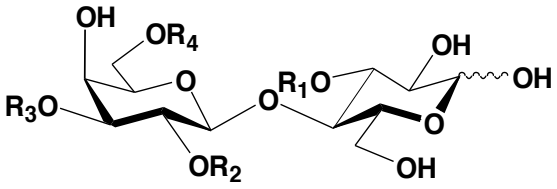


Figure 2.3 ESI mass spectra of aqueous solutions containing (a) 20 μM B1 and 50 μM **L8**, (b) 20 μM B1 and 150 μM **L8**, at pH 7 and 25°C. A P_{ref} (8 μM) was added to each solution to quantify the extent of nonspecific protein-ligand binding during the ESI process. (c) and (d) Normalized distributions of **L8** bound to proteins determined from ESI mass spectra acquired for the solutions described in (a) and (b), respectively.

Listed in Table 2.2 are the results of the ESI-MS binding measurements performed on the twenty-one HMOs. Where binding was detected, the affinities were determined based on at least six measurements performed at multiple ligand concentrations. The errors were reported as the pooled standard deviation. In all cases, the ESI mass spectra were corrected for nonspecific binding using the reference protein method.[32] It should be noted that the reported $K_{a,app}$ values depend on both the intrinsic affinity of each binding site and the number of available binding sites. Because the B1 fragment has a single binding site, the apparent and intrinsic affinities are equivalent. However, A2 possesses two equivalent carbohydrate binding sites.[18] Therefore, the $K_{a,app}$ values are two times larger than the intrinsic affinity.

Table 2.2. Apparent association constants, $K_{a,app}$ (units of 10^2 M^{-1}) for binding of the HMOs (**L1 – 21**) with TcdA-A2 and TcdB-B1 fragments, determined at 25 ºC and pH 7 by the direct ESI-MS assay. [Rows corresponding to parent compound of each series is shown in bold. Error values correspond to one standard deviation. NB (No Binding) no binding detected.]



HMO	R_1	R_2	R_3	R_4	$K_{a,app}$ (A2)	$K_{a,app}$ (B1)
L1	**H**	**H**	**H**	**H**	**NB**	**NB**
L2	H	Fuc	H	H	20 ± 8	12 ± 5
L3	Fuc	H	H	H	NB	NB
L4	Fuc	Fuc	H	H	NB	10 ± 3
L5	H	H	Neu5Ac	H	NB	NB
L6	H	H	H	Neu5Ac	NB	NB
L7	Fuc	H	Neu5Ac	H	7 ± 3	NB
L8	**H**	**H**	**Gal(β1-3)GlcNAc**	**H**	**15 ± 5**	**10 ± 5**
L9	H	H	Fuc(α1-2)Gal(β1-3)GlcNAc	H	8 ± 1	31 ± 2
L10	H	H	Gal(β1-3)[Fuc(α1-4)]GlcNAc	H	7 ± 2	8 ± 4
L11	H	H	Fuc(α1-2)Gal(β1-3)[Fuc(α1-4)]GlcNAc	H	NB	18 ± 9
L12	H	H	Fuc(α1-2)Gal(β1-3)GlcNAc	Gal(β1-4)[Fuc(α1-3)]GlcNAc	NB	21 ± 5
L13	H	H	Neu5Ac(α2-3)Gal(β1-3)GlcNAc	H	NB	NB

L14	H	H	Neu5Ac(α2-6)[Gal(β1-3)]GlcNAc	H	11 ± 2	NB
L15	H	H	Neu5Ac(α2-3)Gal(β1-3)[NeuAc(α2-6)]GlcNAc	H	7 ± 2	NB
L16	H	H	Neu5Ac(α2-3)Gal(β1-3) [Fuc(α1-4)]GlcNAc	H	NB	11 ± 6
L17	H	H	Fuc(α1-2)Gal(β1-3)[NeuAc(α2-6)]GlcNAc	H	NB	NB
L18	**H**	**H**	**Gal(β1-4)GlcNAc**	**H**	**NB**	**15 ± 2**
L19	H	H	Gal(β1-4) [Fuc(α1-3)]GlcNAc	H	17 ± 2	9 ± 4
L20	H	H	Gal(β1-3)[Fuc(α1-4)]GlcNAc(β1-3)Gal(β1-4)[Fuc(α1-3)]GlcNAc	H	NB	NB
L21	H	H	Neu5Ac(α2-6)Gal(β1-4)GlcNAc	H	NB	20 ± 5

The results of the ESI-MS measurements indicate that both the A2 and B1 fragments bind specifically to a number of the HMOs investigated, which range in size from tri- to heptasaccharides. Of the twenty-one HMOs tested, A2 exhibits a measurable affinity for eight of them - five neutral (**L2**, **L8**, **L9**, **L10**, **L19**) and three acidic HMOs (**L7, L14**, **L15**) - while B1 binds to eleven of the HMOs - all of the neutral HMOs, except **L3** and **L20**, and two acidic HMOs (**L16** and **L21**). Neither fragment exhibits a measurable affinity for lactose. However, the trisaccharide **L2**, which contains an additional fucose at the non-reducing end of lactose, binds to both fragments. Interestingly, five of the neutral HMOs (**L2**, **L8**, **L9**, **L10**, and **L19**) are recognized by both A2 and B1. This result points to a degree of structural similarity in the ligand binding sites of the two toxins and raises the possibility of there being common natural human receptors that are recognized by both toxins.

The binding of the toxin fragments to the HMOs is uniformly weak, $K_{a,app} \leq 3100$ M^{-1}, at pH 7 and 25 ^{o}C. The highest affinity ligand for A2 is **L2,** with an $K_{a,app}$ of 2000 M^{-1}, while the most active ligand for B1 is **L9,** with a $K_{a,app}$ of 3100 M^{-1}. However, it is interesting to note that a number of HMOs bind to the toxins with a much higher affinity than the only known natural receptor, α-D-Gal(1,3)β-D-Gal(1,4)β-D-Glc, for which binding constants for both A2 and B1 were found to be ~500 M^{-1}.[18] The results of this study also indicate that increased complexity of the HMOs does not necessarily result in a significant increase in affinity.

The patterns of monovalent, solution-phase ligand binding observed by ESI-MS for A2 show some similarities with the patterns of multivalent, solid-phase ligand binding observed by glycan array screening with the A2 fragment, as well as native

TcdA holotoxin (Consortium for Functional Glycomics (CFG), http:// www.functionalglycomics.org). Most importantly, the Le^X trisaccharide, which is identical to **L2** except for the 2-acetamido group in the residue at the reducing end, is one of the tightest-binding ligands in the glycan array screen for A2 and TcdA holotoxin. Moreover, several more complex oligosaccharides containing Le^X at the reducing end are also some of the tightest binding ligands. The results from the glycan array screening confirm a wide range of modifications can be added at the reducing end of the lactose/LacNAc or Le^X core structures, but most of these modifications add little to increase binding affinity. One of the most interesting differences between the ESI-MS method and glycan array screening, which was also seen in an earlier study,[18] is the ability of the ESI-MS method to measure solution-phase binding constants for ligands of TcdB that have no detectable affinity using the multivalent presentation format of the glycan array. Although the molecular basis for lower affinity multivalent binding in TcdB is not clearly understood at present, differences in the apparent binding affinities of ligands in solution versus multivalent binding interactions with solid-phase glycan arrays have been seen in other systems.[37] As discussed in the earlier ESI-MS study, for example, the three-dimensional arrangement of carbohydrate-binding sites in TcdB may not be compatible with the presentation of ligands in the glycan array, a complication that does not affect single-site measurements peformed using the ESI-MS method.

2.3.2 Docking analysis

To gain more insight into the interaction modalities of HMOs with *C*. *difficile* toxin fragments A2 and B1, molecular docking experiments were performed using Autodock Vina. A standard deviation of about 2-3 kcal mol^{-1} in free energy prediction is typical for current computational docking techniques, including AutoDock Vina.[34,38] Additionally, although severe intra-molecular steric clashes are avoided during docking in Vina, the internal (conformational) free energy is not taken into account during the final ranking of poses. This can lead to significant errors in ranking of the ligand binding conformations (poses), especially when screening flexible molecules like oligosaccharides. In other words, the lowest energy (top ranked) pose is not always consistent with the structure identified experimentally. Several aspects of the docked conformation may be used to filter docking results, such as the presence of key amino acids contacts, structural similarities to known ligands, or the availability of unpaired hydrogen bond donors or acceptors in the protein-ligand complex. Except in rare cases, the conformations of carbohydrate ligands co-crystallized with proteins agree with the exo-anomeric effect of oligosaccaharides. The exo-anomeric effect arises from the overlap of a lone electron pair of the exo-anomeric oxygen with σ* of the O_{ring}-C_1 bond. This favorable stereo-electronic interaction stabilizes conformations, in which lone pair and σ* are properly aligned. Although Autodock Vina has no provisions for the exo-anomeric or other stereo-electronic effects in its scoring function in the free energy evaluations, these conformational constraints can be taken into account either by fixing the Φ dihedral angle (O_{ring}-C_1-O_{exo}-C_i) during molecular docking or by filtering out the putative binding modes that fail to satisfy this rule.

To test the reliability of AutoDock Vina for modeling protein-oligosaccharide interactions, the rabbit receptor of TcdA, α-D-Gal(1,3)β-D-Gal(1,4)β-D-GlcNAc, was prepared using the AMBER force field with Homans' parameters for carbohydrates and docked into the carbohydrate binding site of TcdA-A2 and the results compared to the crystal structure (PDB entry 2G7C). Notably, the selected pose that satisfies the exo-anomeric effect (which was also the lowest energy pose in most of the docking sessions) matches very well (root mean squareue deviation, RMSD <0.1 Å) the pose found in the crystal structure (Figure 2.4). The impressive accuracy of this molecular docking exercise suggests that AutoDock Vina is a suitable computational tool to probe the binding modalities of flexible oligosaccharide ligands, such as the HMOs, to TcdA-A2.

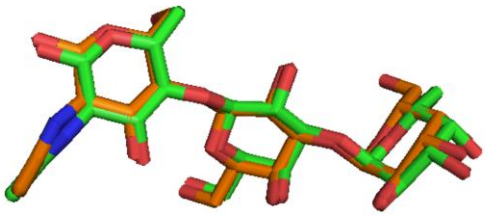


Figure 2.4 PyMol representation of the minimal energy pose (shown as orange sticks) of α-D-Gal(1-3)β-D-Gal(1-4)β-D-GlcNAc obtained following docking with AutoDock Vina and the original pose (shown as green sticks) taken from the crystal structure.

Using this approach, molecular docking was performed on each of the HMOs that were found by ESI-MS to bind to TcdA-A2. Analysis of the docking results reveals that five of the eight ligands (**L2**, **L7**, **L8**, **L9**, and **L10**) share a common binding motif, β-D-GlcNAc(1,3)β-D-Gal(1,4)β-D-Glc, in which the lactose (Lac) disaccharide fragment represents the core recognition element with the lowest RMSD between these ligands (Figure 2.5a). In the case of **L19,** a frame shift was identified, wherein it is the LacNAc fragment that binds, and the Lac moiety does not form contacts with the protein (Figure 2.5b). Notably, this binding motif (involving Lac or LacNAc) matches the position of the LacNAc fragment of the trisaccharide receptor α-D-Gal(1-3)β-D-Gal(1-4)β-D-GlcNAc identified from the crystal structure.[29] In the case of **L14** and **L15**, repeated docking experiments failed to identify likely binding poses (i.e., no poses were obtained in which all Φ dihedral angles were consistent with the exo-anomeric effect). This finding may reflect the extreme flexibility of the glycosidic bond in neuraminic acid.[39] The docking results also provide an explanation for the absence of binding in at least two of the HMOs tested. According to the binding poses shown in Figure 2.5a, the hydroxymethylene group at C6 of β-galactose fits within a distinct indentation in the binding pocket, formed by Glu84, Tyr85, Ile101, Ser121, and Lys122. It is reasonable to expect that any substitutions at this position of Gal will preclude this interaction (or else radically change the mode of binding). The absence of binding observed for **L6** or **L12**, which have sialic acid or Gal(1,4)[Fuc(1,3)]GlcNAc at the C6 position, respectively, can be understood on the basis of this argument.

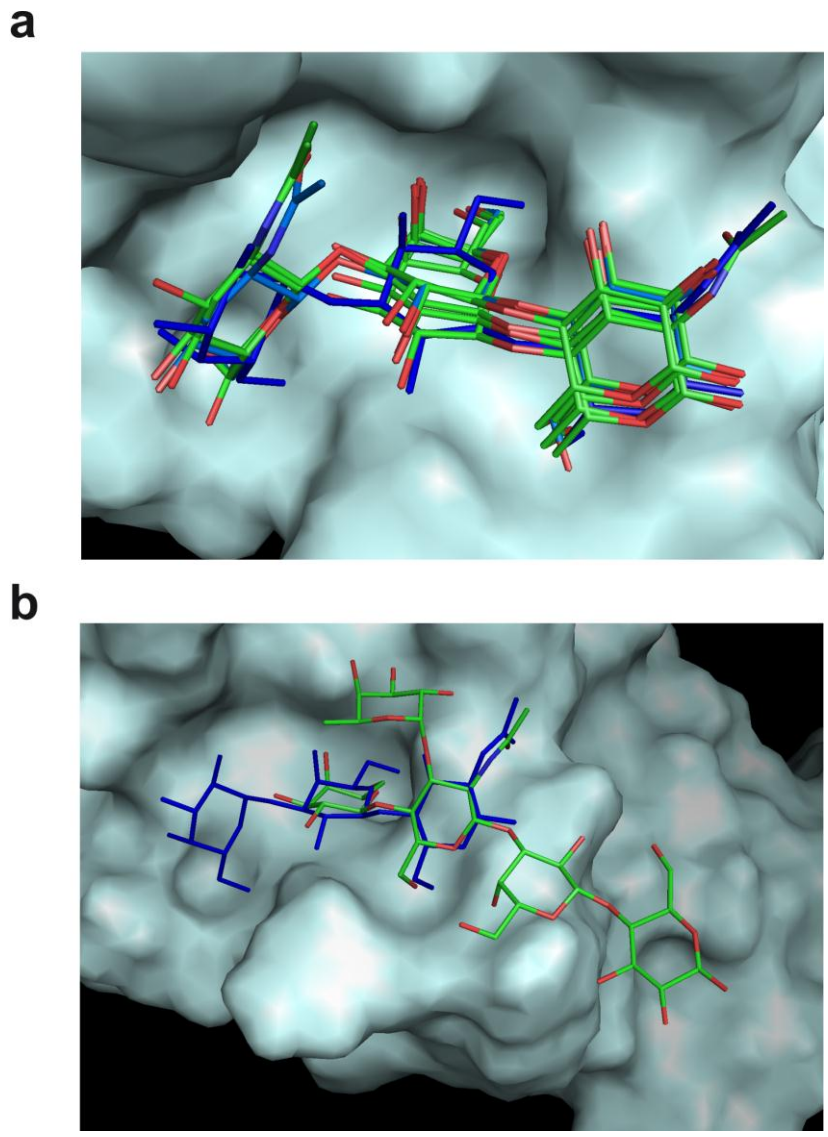
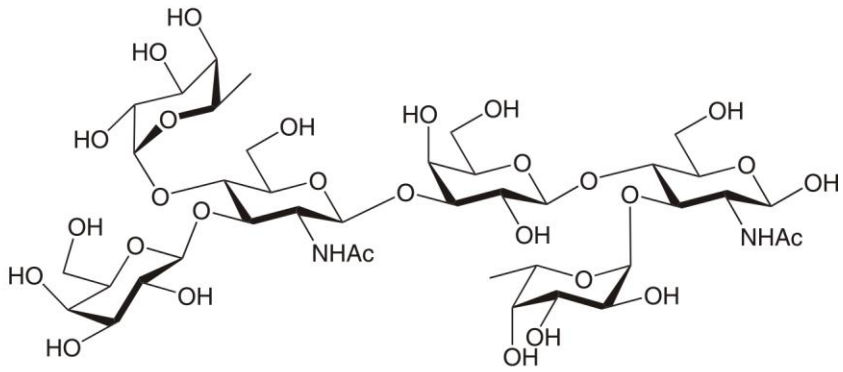


Figure 2.5 (a) Representative poses found for docking of **L2**, **L7**, **L8**, **L9** and **L10** and also shown the position of the trisaccharide α-D-Gal(1-3)β-D-Gal(1-4)β-D-GlcNAc (shown in blue) co-crystallized with TcdA-A2 (PDB entry 2G7C). For clarity, the structures are truncated beyond the trisaccharide GlcNAcβ(1,3) βGal(1,4)βGlc. (b) Frame shift observed in docking of **L19** to TcdA-A2 compared with the reference trisaccharide α-D-Gal(1-3)β-D-Gal(1-4)β-D-GlcNAc (shown in blue).

Molecular docking was also performed to evaluate the interactions between the HMOs and TcdB-B1. Due to the lack of a crystal structure for TcdB-B1, a homology model for TcdB-B1 generated by Modeller[35] was used. In an effort to "refine" the structure of the binding site, geometry optimization was performed on the B1 fragment in the presence of a virtual ligand (**VL**) (Figure 2.6a), which was based on the structures of the HMOs that were found to bind both toxin A and toxin B. To construct **VL**, the molecular structures of all the HMOs that bind both TcdA-A2 and TcdB-B1 (**L2**, **L8**, **L9**, **L10** and **L19**) were considered. Since the purpose of the virtual ligand was to induce TcdB-B1 to adopt the bioactive conformation, it was desirable to maximize the number of intermolecular contacts. To this end **L8**, which effectively fills the binding pocket of TcdA-A2, was chosen as the core structure of **VL**. Two branching α-fucose residues were also added, one linked to the GlcNAc in the third position (to resemble **L10**) and the other fucose linked to the reducing end Glc (to resemble **L19**). The Φ and Ψ dihedral angles in the glycosidic bonds of both attached monosaccharides were set equal to those found in **L10** and **L19**. Additionally, the reducing end Glc was replaced with GlcNAc in order to mimic **L19**. The geometry of **VL** was first optimized using the AMBER force field with Homans' parameters for carbohydrates, after which it was introduced to the putative binding site of the homology model for TcdB-B1. The conformation and binding pose of the resulting energy minimized **VL** structure is shown in Figure 2.6b. Geometry optimization was carried out on the complex of the initial homology model of TcdB-B1 and **VL** with the AMBER force field with Homans' parameters for carbohydrates in a stepwise fashion. First, the protein-ligand interactions were optimized while keeping the protein

coordinates fixed; the constraints on protein structure were then released and the energy of the complex was minimized. Comparison of the initial and optimized structures revealed that the most significant structural changes involved the conformations of amino acid side chains that participate in intermolecular contacts; the overall backbone conformations of both protein and oligosaccharide underwent minimal changes. Also, in the minimized complex, **VL** assumes a conformation in which Φ dihedral angles in all glycosidic bonds are consistent with the exo-anomeric effect.

a



b

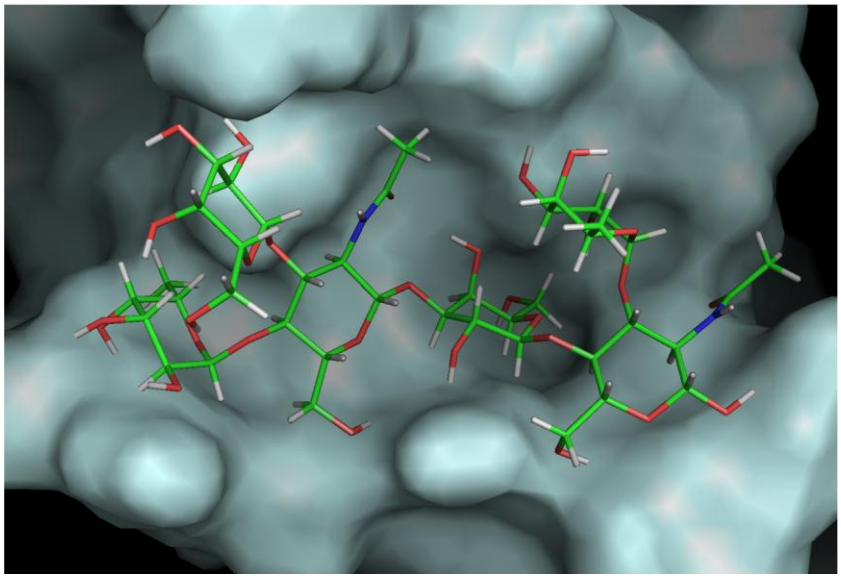


Figure 2.6 (a) Structure of virtual Ligand (**VL**). (b) Structure of **VL** docked in the refined TcdB-B1 binding site.

Docking of eight HMOs (**L2**, **L4**, **L8, L9 L10**, **L18**, **L19** and **L21**) into the refined TcdB-B1 binding site yielded poses that were analogous to the structures found for TcdA-A2 (Figure 2.7). However, attempts to dock the largest HMOs (**L11**, **L12** and **L16**) failed to produce consistent results, presumably due to the inherent flexibility of these oligosaccharides. Notably, the molecular docking results suggest that, despite a number of differences in amino acid sequence between TcdA and TcdB, the general mode of carbohydrate recognition may be conserved. For example, a lactose disaccharide appears to occupy the central portion of the carbohydrate binding site for both toxins.

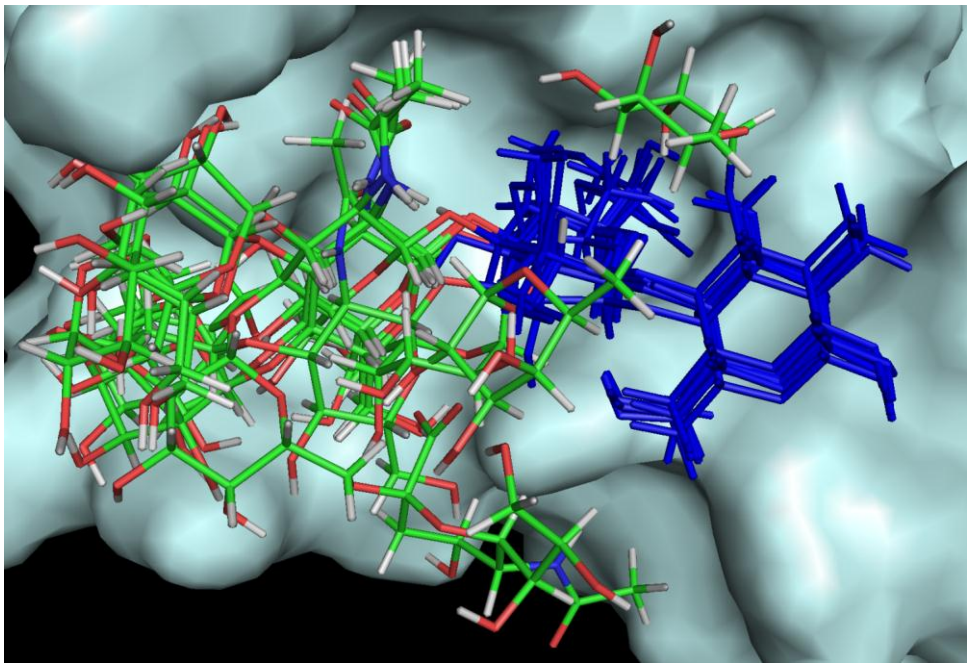


Figure 2.7 Preferred poses of oligosaccharides **L2**, **L4**, **L8**, **L9**, **L10, L18**, **L19** and **L21** docked into refined TcdB-B1 binding site. Lactose fragments are colored in blue.

It should be noted that the use of a homology model for TcdB in this study likely biases the modeled structure towards the template structure and fails to show some of the true structural differences between TcdA and TcdB. In addition, the docking approach used in this study fails to account for water-mediated interactions, which may play important roles in ligand binding. Nevertheless, analysis of some of the docked structures helps to explain how differences in sequence between TcdA and TcdB, especially the different distributions of negatively and positively charged side chains in the binding pocket, can account for some of the observed differences in ligand specificity. For example, the weak binding of **L7** to TcdA-A2 can be explained, at least in part, as the result of unfavorable electrostatic repulsion between negatively charged carboxyl groups (neuraminic acid and Asp183) and a compensatory favorable electrostatic interaction provided by the proximity of the positively charged guanidino group from Arg193. In TcdB-B1, this delicate balance appears to be upset by the substitution of the positively charged side chain to the neutral Asn84 and no binding is observed for **L7** (Figure 2.8).

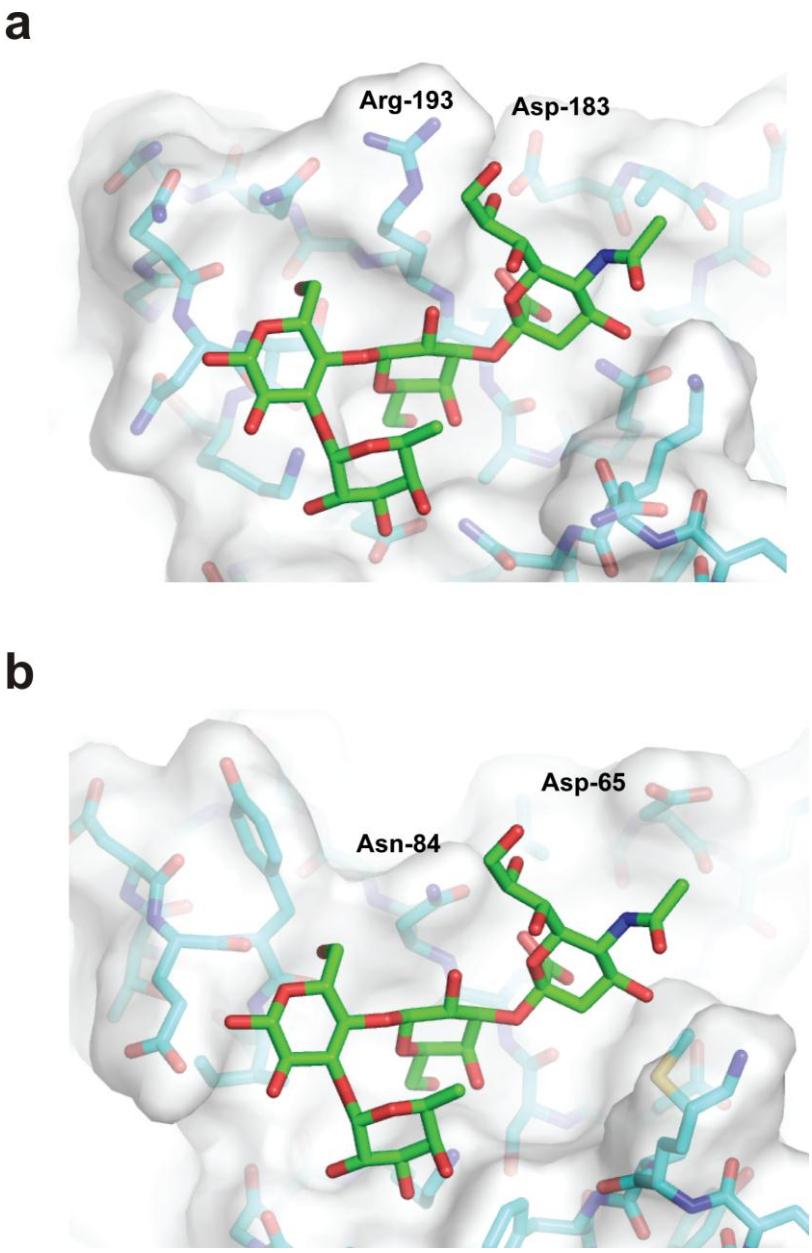


Figure 2.8 Preferred poses of oligosaccharide **L7** docked into carbohydrate-binding sites of (a) TcdA-A2 and (b) TcdB-B1. Spatially equivalent residues that form a salt bridge in TcdA-A2 (Arg-193 and Asp-183) but fail to form a similar interaction in TcdB (Asn-84 and Asp-65) are labeled. A semi-transparent representation of the solvent-accessible surface of the protein is also drawn.

2.3.3 Cytotoxicity neutralization assay

To investigate the inhibitory potential of HMOs on TcdA and TcdB, Vero cytotoxicity neutralization assays were performed using each of the six fractions (I – VI) of HMOs extracted from human milk samples. The results of the assay reveal that none of the HMO fractions inhibited TcdA or TcdB, while the toxin-specific antisera completely neutralized the cytotoxicity of each holotoxin (Figure 2.9). These results are not completely unexpected given that TcdA and, likely, TcdB display linear repeats of multiple carbohydrate binding sites in their carboxy terminal cell binding domains.[30] This arrangement allows the toxin to simultaneously engage multiple glycan receptor sequences on the Vero cell surface, thereby compensating for the observed low affinity interactions between a single glycan receptor sequence and its complimentary carbohydrate binding site. As a consequence, soluble HMOs may not be able to successfully compete with the Vero cells for binding to TcdA or TcdB unless present at very high concentrations. Due to solubility limitations, we were unable to achieve high enough HMO concentrations to demonstrate any possible inhibitory effects.

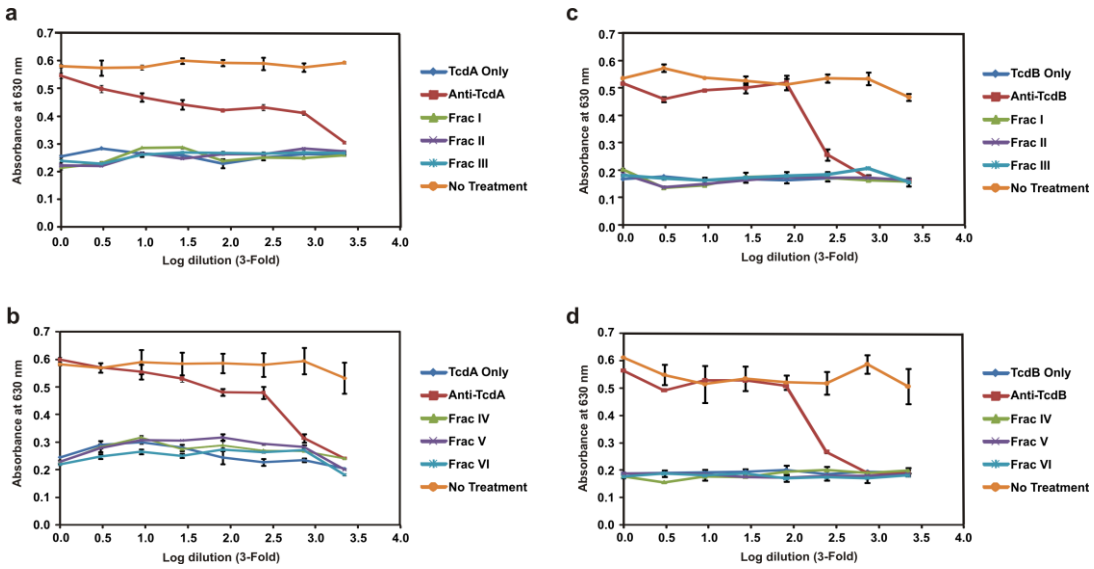


Figure 2.9 HMO Verocytotoxicity neutralization assays. Vero cell monolayers were incubated for 4 h at 37°C with 3 fold serial dilutions of HMO fractions I, II, III, IV, V, and VI or TcdA- (a and b) or TcdB-specific (c and d) rabbit polyclonal antisera.

2.4 Conclusions

Using the direct ESI-MS assay, the binding of fragments of *C*. *difficile* toxins TcdA and TcdB with a library HMOs was investigated. The results of the ESI-MS measurements indicate that both of the toxin fragments investigated, TcdB-B1 and TcdA-A2, bind specifically to HMOs ranging in size from tri- to heptasaccharides. Notably, five of the HMOs tested bind to both toxins- α-D-Fuc(1-2)β-D-Gal(1-4) β-D-Glc, β-D-Gal(1-3)β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc,α-D-Fuc(1-2) β-D-Gal(1-3)β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc, β-D-Gal(1-3)[α-D-Fuc(1-4)]β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc andβ-D-Gal(1-4)[α-D-Fuc(1-3)] β-D-GlcNAc(1-3) β-D-Gal(1-4) β-D-Glc, However, the binding of the HMOs is uniformly weak, with apparent affinities $\leq 10^3$ M^{-1}. The results of molecular docking simulations, taken together with the ESI-MS binding data, suggest that a disaccharide moiety (lactose or lactosamine) represents the core recognition element for both toxin fragments. Verocytotoxicity neutralization assays indicate that the HMOs do not significantly inhibit the cytotoxic effects of TcdA or TcdB. The absence of protection is attributed to the very weak intrinsic affinities that the toxins exhibit towards the HMOs.

2.5 Literature cited

1. Bartlett, J. G. *Clin*. *Infect*. *Dis*. **2008**, *46*, S4-S11.
2. Kelly, C. P.; LaMont, J. T. *Annu*. *Rev*. *Med*. **1998**, *49*, 375-390.
3. Curry, S. R.; Marsh, J. W.; Muto, C. A.; O'Leary, M. M.; Pasculle, A. W.; Harrison, L. H. *J*. *Clin*. *Microbiol*. **2007**, *45*, 215-221.
4. Voth, D. E.; Ballard, J. D. *Clin*. *Microbiol*. *Rev*. **2005**, *18*, 247-263.
5. Lyras, D.; O'Connor, J. R.; Howarth, P. M.; Sambol, S. P.; Carter, G. P.; Phumoonna, T.; Poon, R.; Adams, V.; Vedantam, G.; Johnson, S.; Gerding, D. N.; Rood, J. L. *Nature* **2009**, *458*, 1176-1181.
6. Ballard, J. D. *Nature* **2010**, *467*, 665-666.
7. Kuehne, S. A.; Cartman, S. T.; Heap, J. T.; Kelly, M. L.; Cockayne, A.; Minton, N. P. *Nature* **2010**, *467*, 711-713.
8. Pruitt, R. N.; Chambers, M. G.; Ng, K. K.; Ohi, M. D.; Lacy, D. B. *Proc*. *Natl*. *Acad*. *Sci*. *U S A* **2010**, *107*, 13467-13472.
9. Dove, C. H.; Wang, S. Z.; Price, S. B.; Phelps, C. J.; Lyerly, D. M.; Wilkins, T. D.; Johnson, J. L. *Infect*. *Immun*. **1990**, *58*, 480-488.
10. Voneichelstreiber, C.; Sauerborn, M.; Kuramitsu, H. K. *J*. *Bacteriol*. **1992**, *174*, 6707-6710.
11. Voneichelstreiber, C.; Boquet, P.; Sauerborn, M.; Thelestam, M. *TIM*. **1996**, *4*, 375-382.
12. Rupnik, M.; Wilcox, M. H.; Gerding, D. N. *Nat*. *Rev*. *Microbiol*. **2009**, *7*, 526-536.
13. Gerding, D. N.; Johnson, S. *Clin*. *Infect*. *Dis*. **2010**, *51*, 1306–1313.

14. Heerze, L. D.; Kelm, M. A.; Talbot, J. A.; Armstrong, G. D. *J*. *Infect*. *Dis*. **1994**, *169*, 1291-1296.

15. Dallas, S. D.; Rolfe, R. D. *J*. *Med*. *Microbiol*. **1998**, *47*, 879-888.

16. Krivan, H. C.; Clark, G. F.; Smith, D. F.; Wilkins, T. D. *Infect*. *Immun*. **1986**, *53*, 573-581.

17. Rolfe, R. D.; Song, W. *J*. *Med*. *Microbiol*. **1995**, *42*, 10-19.

18. Dingle, T.; Wee, S.; Mulvey, G. L.; Greco, A.; Kitova, E. N.; Sun, J. X.; Lin, S. J.; Klassen, J. S.; Palcic, M. M.; Ng, K. K. S.; Armstrong, G. D. *Glycobiology* **2008**, *18*, 698-706.

19. Smith, J. A.; Cooke, D. L.; Hyde, S.; Borriello, S. P.; Long, R. G. *J*. *Med*. *Microbiol*. **1997**, *46*, 953-958.

20. Tucker, K. D.; Wilkins, T. D. *Infect*. *Immun*. **1991**, *59*, 73-78.

21. Newburg, D. S. *J*. *Anim Sci*. **2009**, *87*, 26-34.

22. Andersson, B.; Porras, O.; Hanson, B.; Lagergard, T.; Svanborgeden, C. *J*. *Infect*. *Dis*. **1986**, *153*, 232-237.

23. Newburg, D. S.; Ruiz-Palacios, G. M.; Morrow, A. L. *Annu*. *Rev*. *Nutr*. **2005**, *25*, 37-58.

24. Cravioto, A.; Tello, A.; Villafan, H.; Ruiz, J.; Delvedovo, S.; Neeser, J. R. *J*. *Infect*. *Dis*. **1991**, *163*, 1247-1255.

25. Ruiz-Palacios, G. M.; Cervantes, L. E.; Ramos, P.; Chavez-Munguia, B.; Newburg, D. S. *J*. *Biol*. *Chem*. **2003**, *278*, 14112-14120.

26. Idota, T.; Kawakami, H.; Murakami, Y.; Sugawara, M. *Biosci*. *Biotechnol*. *Biochem*. **1995**, *59*, 417-419.

27. Martin-Sosa, S.; Martin, M. J.; Hueso, P. *J*. *Nutr*. **2002**, *132*, 3067-3072.

28. Ninonuevo, M. R.; Park, Y.; Yin, H. F.; Zhang, J. H.; Ward, R. E.; Clowers, B. H.; German, J. B.; Freeman, S. L.; Killen, K.; Grimm, R.; Lebrilla, C. B. *J*. *Agric*. *Food Chem*. **2006**, *54*, 7471-7480.

29. Greco, A.; Ho, J. G. S.; Lin, S. J.; Pacilic, M. M.; Rupnik, M.; Ng, K. K. S. *Nat*. *Struct*. *Mol*. *Biol*. **2006**, *13*, 460-461.

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

31. Ward, R. E.; Ninonuevo, M.; Mills, D. A.; Lebrilla, C. B.; German, J. B. *Appl*. *Environ*. *Microbiol*. **2006**, *72*, 4497-4499.

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

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

34. Trott, O.; Olson, A. J. *J*.*Comput*.*Chem*. **2009**, *31*, 455-461.

35. Fiser, A.; Sali, A. *Methods Enzymol*. **2003**, *374*, 461-491.

36. Homans, S. W. *Biochemistry* **1990**, *29*, 9110-9118.

37. Oyelaran, O.; Gildersleeve, J. C. *Curr. Opin. in chem. Biol.* **2009**, *13*, 406-413.

38. Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. *J*. *Comput*. *Chem*. **2006**, *28*, 1145-52.

39. Bitzer, R. S.; Barbosa, A. G. H.; da Silva, C. O.; Nascimento, M. A. C. *Carbohydr*. *Res*. **2005**, *340*, 2171-2184.

Chapter 3

Quantifying Labile Protein-Ligand Interactions using Electrospray Ionization Mass Spectrometry[*,+]

3.1 Introduction

The direct electrospray ionization mass spectrometry (ESI-MS) assay has emerged as a powerful technique for quantifying binding constants (K_a) for protein-ligand interactions, as well as other noncovalent biological complexes, in solution.[1-3] The assay is based on the direct detection and quantification of the abundance (*Ab*) of ligand-bound and unbound protein ions in the gas phase, e.g., PL^{n+} and P^{n+}, respectively. A key assumption is that the measured abundance ratio (*R*) is equivalent to the equilibrium concentration ratio of ligand-bound and free protein in solution, eq 1.3. From the measured *R* value and initial concentrations of protein ($[P]_o$) and ligand ($[L]_o$), K_a can be calculated,[4] using eq 1.2.

The direct ESI-MS assay has been used to measure affinities for a range of protein-ligand complexes, including antibody-antigen, lectin-carbohydrate, enzyme-substrate/inhibitor complexes, and in many instances the K_a values agree well with constants obtained by other analytical methods, including isothermal titration calorimetry (ITC), surface plasmon resonance and frontal affinity chromatography MS.[4-10] However, there have also been reports of protein-ligand complexes that could

---

[*] A version of this chapter has been published: El-Hawiet, A.; Kitova, E. N.; Liu, L.; Klassen, J. S. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1893-1899.

[+] Se155-4 scFv expression and purification were done by Blake Zheng (University of Alberta)

not be detected by ESI-MS or, if detected, the relative abundance of ligand-bound and unbound protein ions did not match the distribution expected in solution, with less binding observed in the gas phase.[11-14] These anomalous results are often due to the occurrence of in-source dissociation, whereby the gaseous complexes undergo partial or complete dissociation during ESI-MS analysis. If the gas phase PL ions are kinetically labile and undergo dissociation during analysis, the magnitude of the measured *R* value and, correspondingly, the K_a value will be artificially low. In the extreme case, where no PL ions survive, in-source dissociation will result in a false negative. Recently, it was shown that solution or gas phase additives can, in some instances, protect complexes from in-source dissociation.[12,15] However, this approach does have its limitations and the detection of very labile gas phase complexes, which rapidly dissociate at ambient temperature, by ESI-MS remains problematic.

Here, we describe an indirect ESI-MS approach to quantify protein-ligand interactions that are highly labile and prone to in-source dissociation. The method, referred to as the *reference ligand* ESI-MS method, employs direct ESI-MS analysis in conjunction with a reference ligand (L_{ref}). The L_{ref} binds specifically to P, at the same binding site as L, with known affinity and forms a stable protein-ligand complex in the gas phase. The fraction of P bound to L_{ref}, which is determined directly from the ESI mass spectrum, is sensitive to the fraction of P bound to L in solution and enables the affinity of the PL complex to be established. A mathematical framework for the implementation of the method in cases where P has one or two specific ligand binding sites is given. To demonstrate the reliability of the method, monosaccharide affinities were measured for two carbohydrate-binding proteins, a single chain fragment of a

monoclonal antibody binding and the lectin concanavalin A, and the values compared with data obtained using ITC.

3.2 Experimental Methods

3.2.1 Proteins and ligands

The carbohydrate-binding antibody single chain fragment, Se155-4 scFv (MW 26 539 Da), was produced using recombinant technology.[16] The scFv was concentrated and dialyzed against aqueous solution of 50 mM ammonium acetate (pH 7) using microconcentrators (Millipore Corp., Bedford, MA) with a molecular weight cut-off of 10 kDa, and lyophilized prior to MS analysis. The scFv was weighed immediately after removing it from the lyophilizer, dissolved in a known volume of aqueous 50 mM ammonium acetate and stored at –20 ^{o}C if not used immediately. The proteins concanavalin A (ConA, MW 25 600 Da for monomer) and lysozyme (Lyz, MW 14 311 Da) were purchased from Sigma-Aldrich Canada (Oakville, Canada), and used without further purification. The synthetic carbohydrate ligands, Methyl-α-D-Abe (**1**), Methyl-α-D-3-deoxy Ara (**2**), Methyl-α-D-Glc (**3**) and Methylα-D-Abe-(1-3)-2-*O*-methyl-α-D-Man-(1-3)-α-D-Glc-(1-4)-β-D-Glc (**4**) were provided by Prof. D. Bundle (University of Alberta), while Methyl-α-D-Man (**5**) and Methyl-α-D-3,6-di-O-(α-D-Man)-α-D-Man (**6**) were purchased from Toronto Research Chemicals Inc (North York, Canada).

3.2.2 Mass spectrometry

All experiments were performed on an Apex II 9.4 tesla Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometers (Bruker, Billerica, MA) equipped with an external nanoflow ES ion source. A description of the instrument and the experimental and instrumental parameters used in the ESI-MS binding measurements is given elsewhere[4] and described in Chapter 1.

3.3 Results and discussion

3.3.1 Overview of reference ligand method

The influence of in-source dissociation on ESI-MS mass spectra acquired for a solution of P and two ligands, L and L_{ref}, which bind at the same site, is depicted in Figure 3.1. In the absence of in-source dissociation (and assuming uniform response factors and an absence of nonspecific ligand binding), the measured abundance ratios of ligand-bound to free P ions (R and R_{ref}) will reflect the concentration ratios in solution (Figures 3.1a,b) and K_a for the PL complex can be calculated using eq 3.1:

$$K_a = \frac{R}{[L]_o - \frac{R}{1+R+R_{ref}}[P]_o} \tag{3.1}$$

where R is given by eq 1 and R_{ref} by eq 3.2:

$$\frac{[PL_{ref}]_{eq}}{[P]_{eq}} = \frac{\sum_n Ab(PL_{ref}{}^{n+})}{\sum_n Ab(P^{n+})} = R_{ref} \tag{3.2}$$

If, on the other hand, the gas phase ions of the PL complex are kinetically labile and undergo in-source dissociation, the relative abundance of ligand-bound and free protein ions measured by ESI-MS will no longer reflect the original concentration ratios (Figures 3.1a,c). In the extreme case, where no PL ions survive until detection,

the measured abundance ratio of L_{ref}-bound P to free P gas phase ions (i.e., $R_{ref,app}$) will depend on the concentration of both free P and PL originally present in solution, eq 3.3:

$$\frac{[\mathrm{PL_{ref}}]_{eq}}{[\mathrm{P}]_{eq}+[\mathrm{PL}]_{eq}} = \frac{\sum_n Ab(\mathrm{PL_{ref}}^{n+})}{\sum_n Ab(\mathrm{P}^{n+})} = R_{ref,app} \tag{3.3}$$

However, if $K_{a,ref}$ is known, R_{ref} (the "true" concentration ratio $[PL_{ref}]/[P]$) can be calculated from $R_{ref,app}$, eq 3.4:

$$R_{ref} = \mathrm{K_{a,ref}}([\mathrm{L_{ref}}]_o - [\mathrm{P}]_o \frac{R_{ref,app}}{R_{ref,app}+1}) \tag{3.4}$$

Furthermore, the magnitude of R can be calculated from the corresponding $R_{ref,app}$ and R_{ref} values, eq 3.5:

$$R = \frac{R_{ref}}{R_{ref,app}} - 1 \tag{3.5}$$

Once R and R_{ref} are known, the K_a value for the PL complex can be calculated using eq 3.1.

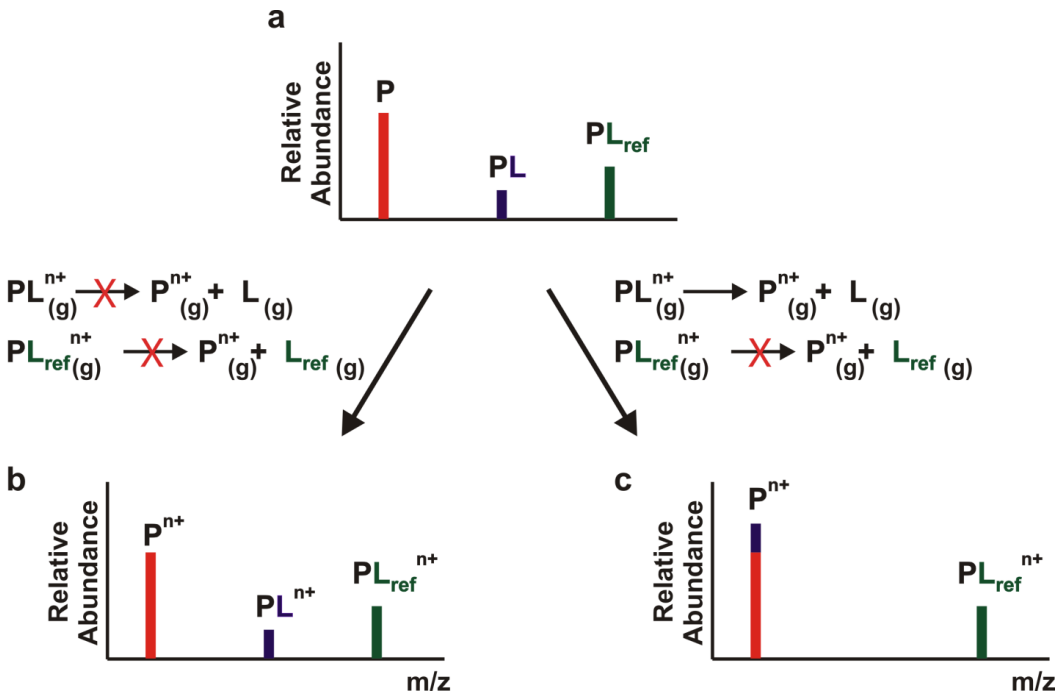


Figure 3.1 Influence of in-source dissociation on the relative abundance of gas phase protein-ligand complexes. (a) Hypothetical distribution of species present in solution when ligands L and L_{ref} interact with a single binding site of protein P. (b) In the absence of in-source dissociation, the relative abundance of gas phase PL^{n+} and PL_{ref}^{n+} ions will reflect the original distribution in solution. (c) In-source dissociation of the PL^{n+} ions (but not the PL_{ref}^{n+} ions) will result in a decrease in the measured abundance ratio of PL_{ref}^{n+} to P^{n+} ions.

It should be noted that the above treatment can easily be extended to the case where the gas phase PL ions undergo partial dissociation in the source. In this situation, R_{ref} can be determined from the measured abundance ratio of L_{ref}-bound P to free P gas phase ions (i.e., $R_{ref,app}$) and the measured abundance ratio of L-bound P to free P gas phase ions (i.e., R_{app}) using eq 3.6:

$$R_{ref} = \mathrm{K}_{\mathrm{a,ref}}([\mathrm{L}_{\mathrm{ref}}]_{\mathrm{o}} - [\mathrm{P}]_{\mathrm{o}} \frac{R_{ref,app}}{R_{ref,app} + R_{ref,app} + 1}) \tag{3.6}$$

and R can be found using eq 3.7:

$$R = \frac{(1 + R_{app})}{R_{ref,app}} R_{ref} - 1 \tag{3.7}$$

Once R and R_{ref} are known, K_a can be calculated in the usual way.

Because many proteins possess multiple ligand binding sites, it is also useful to consider how the *reference ligand* ESI-MS method can be applied to these situations. Given below are the relevant mathematical expressions for the application of the method for the case where P possesses two equivalent and independent ligand binding sites. Expressions for the equilibrium constants ($K_{a,1}$ and $K_{a,2}$) for the sequential binding of L to P in the presence of L_{ref}, written in terms of the concentration ratios of ligand bound (L or L_{ref}) P to free P, are:

$$\mathrm{K}_{a,1} = \frac{R_{1,0}}{[\mathrm{L}]_{\circ} - \frac{(R_{1,0} + 2R_{2,0} + R_{1,1})[\mathrm{P}]_{\mathrm{o}}}{1 + R_{1,0} + R_{2,0} + R_{0,1} + R_{1,1} + R_{0,2}}} \tag{3.8a}$$

$$\mathrm{K}_{a,2} = \frac{R_{2,0}}{R_{1,0}\left([\mathrm{L}]_{\circ} - \frac{(R_{1,0} + 2R_{2,0} + R_{1,1})[\mathrm{P}]_{\circ}}{1 + R_{1,0} + R_{2,0} + R_{0,1} + R_{1,1} + R_{0,2}}\right)} \tag{3.8b}$$

where $R_{i,j}$ is the concentration ratio of ligand-bound P (to i molecules of L and j molecules of L_{ref}) to free P. As before, in the absence of in-source dissociation, the $R_{i,j}$ values can be determined directly from the relative abundance of ligand-bound and free P ions measured by ESI-MS, eq 3.9:

$$\frac{[\mathrm{P(L)}_i(\mathrm{L}_{\mathrm{ref}})_j]_{eq}}{[\mathrm{P}]_{eq}} = \frac{\sum_{\mathrm{n}} Ab(\mathrm{PL}_i\mathrm{L}_{\mathrm{ref},j}^{\mathrm{n+}})}{\sum_{\mathrm{n}} Ab(\mathrm{P}^{\mathrm{n+}})} = R_{i,j} \tag{3.9}$$

However, if in-source dissociation takes place, the measured ratios will no longer reflect solution composition. In the extreme case, where none of the P-L interactions survive the ion source, only the ratios $R_{0,1app}$ and $R_{0,2app}$ can be determined from the ESI mass spectrum:

$$\frac{\sum_{\mathrm{n}} Ab(\mathrm{PL}_{\mathrm{ref}}^{\mathrm{n+}})}{\sum_{\mathrm{n}} Ab(\mathrm{P}^{\mathrm{n+}})} = R_{0,1app} \tag{3.10a}$$

$$\frac{\sum_{\mathrm{n}} Ab(\mathrm{P(L}_{\mathrm{ref}})_2^{\mathrm{n+}})}{\sum_{\mathrm{n}} Ab(\mathrm{P}^{\mathrm{n+}})} = R_{0,2app} \tag{3.10b}$$

Using the known values of $K_{a1,Lref}$, $[L_{ref}]_o$, $[P]_o$, $R_{0,1app}$ and $R_{0,2app}$, the $R_{i,j}$ terms can be calculated using following expressions:

$$R_{0,1} = \mathrm{K}_{\mathrm{a1,Lref}}([\mathrm{L}_{\mathrm{ref}}]_{\mathrm{o}} - [\mathrm{P}]_{\mathrm{o}} \frac{R_{0,1app} + 2R_{0,2app}}{1 + R_{0,1app} + R_{0,2app}}) \tag{3.11a}$$

$$R_{0,2} = (R_{0,1})^2 / 4 \tag{3.11b}$$

$$R_{1,1} = \frac{R_{0,1app}R_{0,2}}{R_{0,2app}} - R_{0,1} \tag{3.11c}$$

$$R_{1,0} = -2 + 2\sqrt{\frac{R_{0,2}}{R_{0,2app}}} \quad (3.11d)$$

$$R_{2,0} = 1 + \frac{R_{0,2}}{R_{0,2app}} - 2\sqrt{\frac{R_{0,2}}{R_{0,2app}}} \quad (3.11e)$$

and $K_{a,1}$ and $K_{a,2}$ can be calculated using eqs 3.8a and 3.8b.

The two basic requirements for a suitable L_{ref} are that it binds specifically to P in solution, at the same binding site as L, with a known affinity and that it forms a stable protein-ligand interaction in the gas phase, i.e., is resistant to in-source dissociation. In addition there are several other practical considerations. It is desirable to use a reasonably strong binding ligand as L_{ref}, with a $K_{a,ref}$ of $10^5 - 10^7$ M^{-1}, since in this case R_{ref} and $R_{ref,app}$ are more sensitive to the presence of the competing ligand. Additionally, changes in R_{ref} are more pronounced when low protein concentrations (~μM) are used. Finally, depending on the magnitude of K_a and $K_{a,ref}$, the concentrations of P and both ligands (L and L_{ref}) may need to be adjusted so that both complexes are present in solution at significant concentrations.

3.3.2 Determination of protein-ligand affinities

To demonstrate the reliability of the *reference ligand* ESI-MS method for quantifying protein-ligand interactions that are not readily detected by ESI-MS, binding measurements were carried out on two carbohydrate-binding proteins, Se155-4 scFv and ConA. Association constants have been determined by ITC for the binding of Se155-4 scFv to the monosaccharides **1** (1.5 x 10^3 M^{-1}) and **2** (1.2 x 10^2 M^{-1}) at pH 7 and 25 ^oC.[17] The monosaccharide **3**, which was shown by ITC not to bind to the scFv,[18] served as a negative control. Con A is a tetramer above pH 7 and a dimer

below pH 6. Each subunit possesses a single carbohydrate binding site with specificity for the α-pyranose forms of Glc or Man. The affinity of dimeric ConA for the monosaccharide **5** was determined to be (7.9 ± 1.0) x 10^3 M^{-1} at pH 5.2 and 25 ^oC by ITC.[19]

a. scFv-monosaccharide binding

Shown in Figure 3.2a is a representative ESI mass spectrum acquired for a solution of scFv (10 μM) and **1** (1 mM) at pH 7 and 25 ^oC. A reference protein (Lyz) was also added to monitor for the occurrence of nonspecific ligand binding to scFv during the ESI process.[20] At these concentrations, 60% of the scFv is expected to be bound to **1** in solution. However, no gas phase ions corresponding to the specific (scFv + **1**) complex were detected, indicating the occurrence of in-source dissociation. Attempts to stabilize the complex during ESI-MS analysis using a high concentration (10 mM) of imidazole, a stabilizing additive,[12] were unsuccessful (data not shown). The instability of the gas phase ions of the (scFv + **1**) complex can be explained by the small number of intermolecular hydrogen bonds that **1** is capable of making. Direct ESI-MS analysis of solutions of scFv (10 μM) with **2** (2 mM) or **3** (1 mM) also failed to detect ions of the (scFv + **2**) or (scFv + **3**) complex (data not shown). In contrast, ESI-MS analysis of a solution of scFv (10 μM) and **4** (59 μM) clearly identified the presence of gas phase ions corresponding to the (scFv + **4**) complex, Figure 3.2b. The K_a value determined directly from the ESI mass spectrum, following correction for nonspecific ligand binding,[20] was (1.6 ± 0.5) x 10^5 M^{-1}. Shown in Figure 3.2c is a representative ESI mass spectrum acquired for a solution of scFv (10 μM), **1** (1 mM) and **4** (59 μM). Although ions corresponding to the (scFv + **1**) complex were not detected, the addition

of **1** to the solution resulted in a decrease in the fraction of scFv bound to **4**, indicating the presence of specific binding between scFv and **1** in solution. Plotted in Figure 3.3 is the fraction of bound (to **4**) and unbound scFv determined by direct ESI-MS measurements at fixed concentrations of scFv (10 μM) and **4** (59 μM) and varying concentrations of **1** (0, 500, 1000 and 2000 μM). Notably, the fraction of scFv bound to **4** decreased with increasing concentration of **1**. Analysis of the ESI-MS data using the approach described above leads to an average K_a value of (1.4 ± 0.3) x 10^3 M^{-1} for the (scFv + **1**) complex, Table 3.1. This value is indistinguishable, within experimental error, from the ITC-derived value, of (1.5 ± 0.4) x 10^3 M^{-1}.[17] Following the same approach, a K_a value of (1.7 ± 0.9) x 10^2 M^{-1} was determined for the (scFv + **2**) complex, which is also in good agreement with the ITC value of (1.2 ± 0.5) x 10^2 M^{-1}.[17] In contrast, the assay did not detect any binding between the scFv and **3**, consistent with the results of ITC measurements.[18]

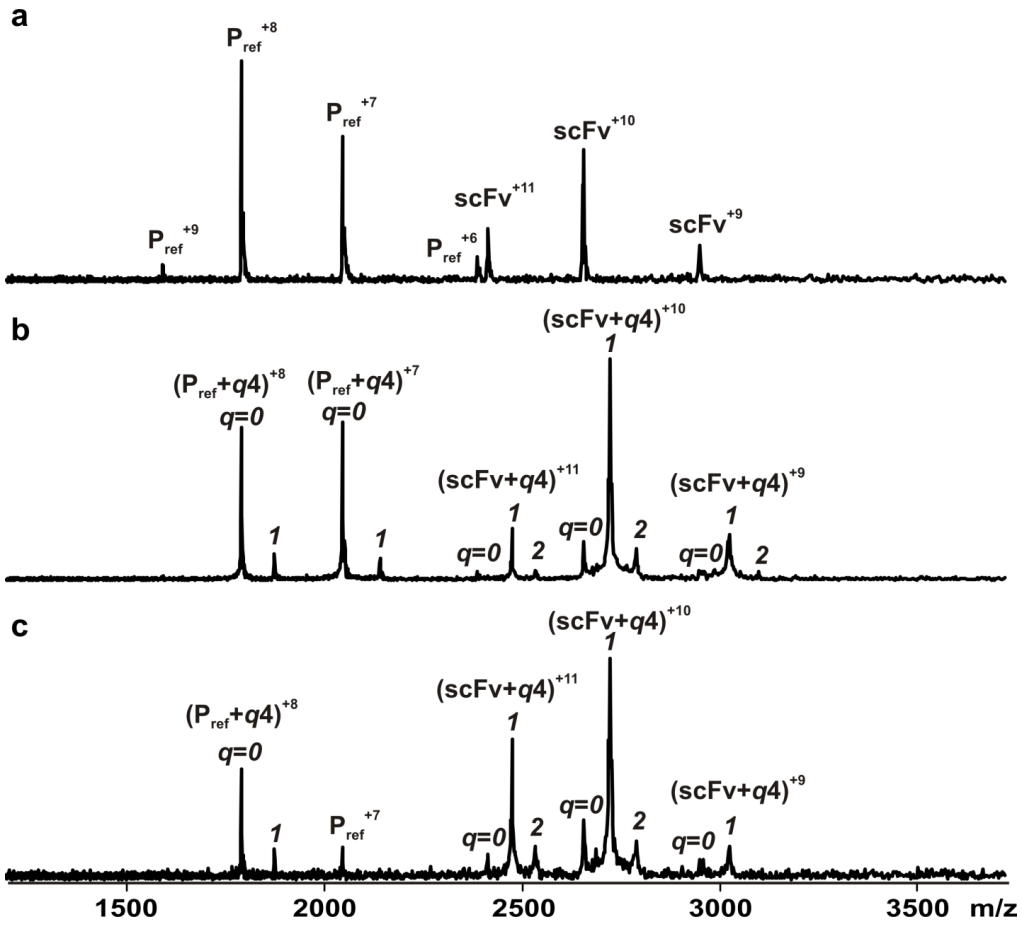


Figure 3.2 ESI mass spectra obtained for the solutions of Se155-4 scFv (10 μM) and (a) **1** (1000 μM), (b) **4** (59 μM) and (c) **4** (59 μM) and **1** (1000 μM). The number of molecules of **4** bound to the protein ions is indicated by *q*.

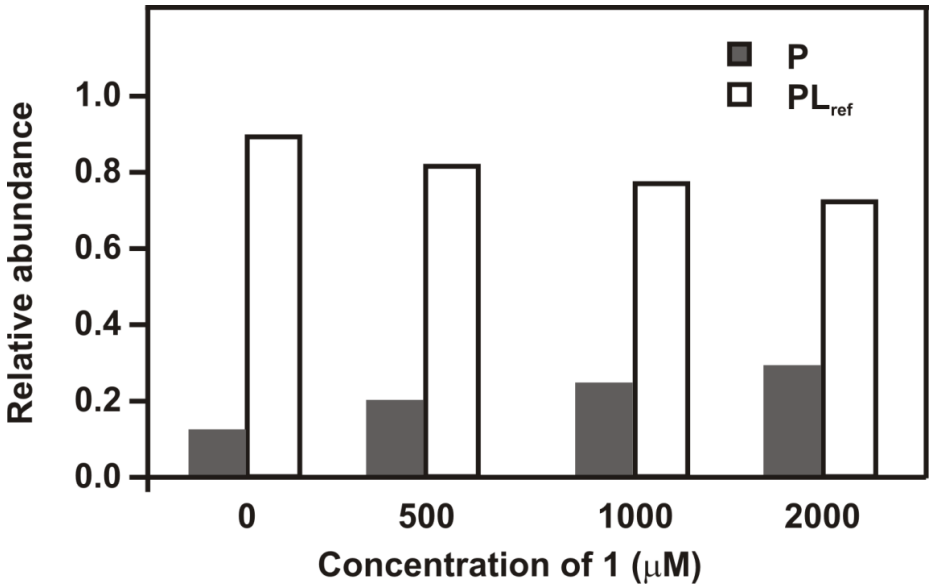


Figure 3.3 Distribution of the relative abundance of scFv (≡ P) and (scFv + **4**) complex (≡ PL) measured by ESI-MS for solutions of scFv (10 μM), **4** (59 μM) and **1** at concentrations of 0, 500, 1000 and 2000 μM.

Table 3.1 Association constants (K_a) for carbohydrate ligand binding to Se155-4 scFv and ConA determined by the ESI-MS *reference ligand method* and by ITC at 25 °C.[a]

Protein	Ligand	pH	ESI-MS K_a x 10^{-3} (M^{-1})	ITC K_a x 10^{-3} (M^{-1})
scFv	**1**	7.0	1.4 ± 0.3	1.5 ± 0.4 [b]
scFv	**2**	7.0	0.17 ± 0.09	0.12 ± 0.05 [b]
scFv	**3**	7.0	NB [c]	NB [c]
ConA	**5**	5.2	10 ± 4 [d]	7.9 ± 1.0 [d,e]

a. Errors correspond to one standard deviation. b. Values taken from reference 16. c. NB ≡ No binding detected. d. Values correspond to average intrinsic K_a for dimeric ConA. e. Value taken from reference 19.

b. ConA-monosaccharide binding

To demonstrate that the *reference ligand* ESI-MS method is generally applicable to the quantification of labile protein-ligand interactions, the method was also used to measure the affinity of the lectin ConA for the monosaccharide **5**. The binding measurements were performed at pH 5.2, where ConA exists predominantly as a homodimer. At this pH, ConA is known to bind the monosaccharide **5**, as well as oligomannose ligands such as **6**, which served as L_{ref} for these measurements.[21] Importantly, it was previously shown that **5** and **6** bind in the same binding site.[22] In an earlier study, it was reported that the ConA-**5** interaction is quite labile in the gas phase and can be, depending on the ESI-MS instrumentation used, difficult to detect.[11] In the present work, it was found that the complex could be directly detected by ESI-MS. However, the relative abundance of the ligand-bound forms of ConA dimer was sensitive to ion source conditions, such as hexapole accumulation time. In order to demonstrate the reliability of the assay in cases where the complex could not be directly detected, source conditions that lead to complete dissociation of the ConA-**5** interactions were used.

Shown in Figure 3.4a is an ESI mass spectrum measured for a solution of ConA (30 μM for monomer) at pH 5.2. Notably, only ions corresponding to ConA homodimer ($ConA_2$) were identified; no ions corresponding to monomer or homotetramer were detected. Upon addition of **6** (15 μM) to the solution, ions corresponding to $ConA_2$ bound to one and two molecules of **6** were also detected (Figure 3.4b). The intrinsic affinity of $ConA_2$ for **6**, based on the $K_{a,1}$ and $K_{a,2}$ values determined directly from the ESI mass spectra, is $(5 \pm 2) \times 10^5$ M^{-1}. This value is in

excellent agreement with a value of (5.1 ± 0.2) x 10^5 M^{-1} that was determined by ITC for homodimer of ConA at pH 5.2.[21] Under the experimental conditions used, ESI-MS analysis of a solution of $ConA_2$ (15 μM) with **5** (50 or 100 μM) and **6** (15 μM) led to the detection of ions corresponding to $ConA_2$ bound to one and two molecules of **6** but not ions corresponding to $ConA_2$ bound to **5** (Figure 3.4c). However, the addition of **5** to the solution did result in a small but measurable reduction in the fraction of $ConA_2$ bound to **6**. Following the procedures outlined above, the intrinsic K_a for the $ConA_2$**-5** interaction was determined to be (1.0 ± 0.4) x 10^4 M^{-1}, which is in reasonable agreement with the ITC-derived value of (7.9 ± 1.0) x 10^3 M^{-1}, Table 3.1.[19]

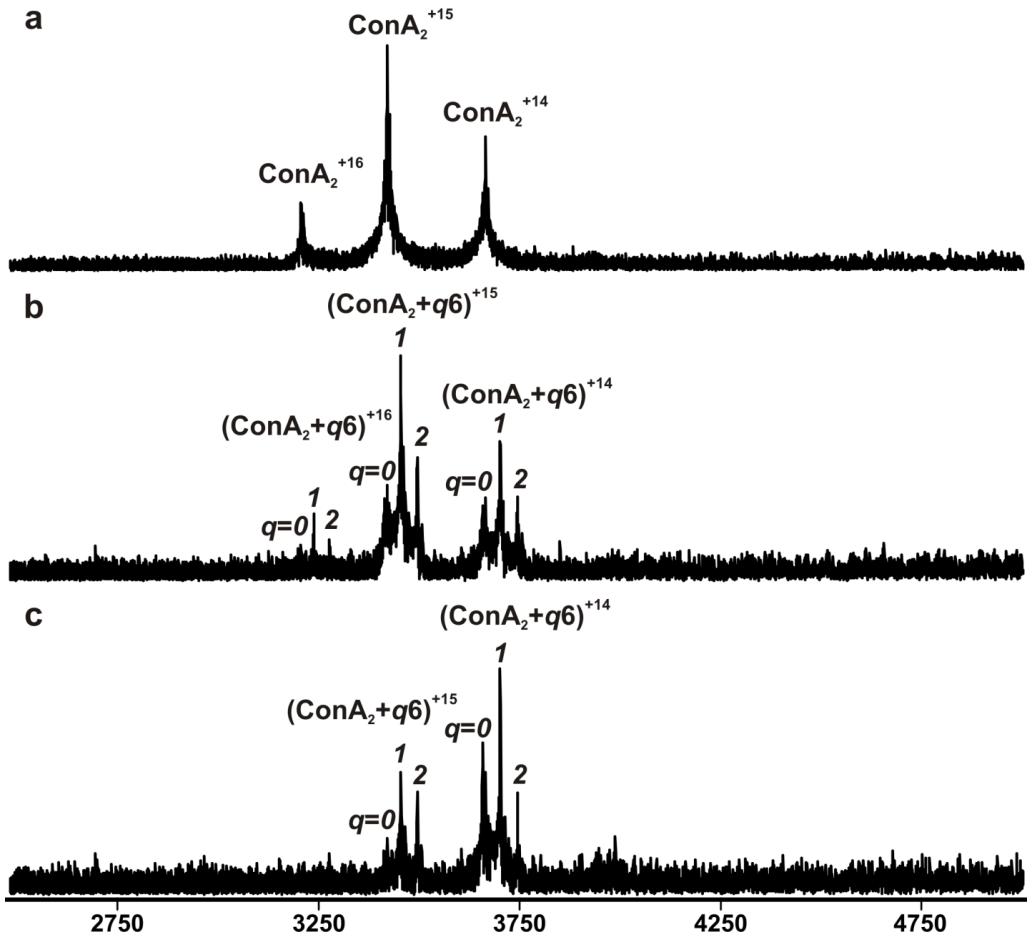


Figure 3.4 ESI mass spectra obtained for the solutions of (a) $ConA_2$ (15 μM) alone and with (b) **6** (15 μM) and (c) **6** (15 μM) and **5** (100 μM). The number of molecules of **6** bound to the protein ions is indicated by *q*.

3.4 Conclusions

In summary, an ESI-MS approach for quantifying protein-ligand interactions that are prone to in-source dissociation and, therefore, difficult to detect directly is described. The *reference ligand* ESI-MS method employs the direct ESI-MS assay in conjunction with a L_{ref}, which binds competitively to protein of interest with known affinity and forms a stable complex in the gas phase. The relative abundance of L_{ref}-bound protein to free protein, which can be measured directly by ESI-MS, is sensitive to the presence of other ligands in solution that compete for the same binding site. As a result, it is possible to quantify protein-ligand interactions that are unstable in the gas phase by measuring the relative abundance of L_{ref}-bound protein using the direct ESI-MS assay. The relevant mathematical expressions for the implementation of the method for proteins with a single ligand binding site or two equivalent binding sites are given. To demonstrate the reliability of the method, the binding of monosaccharide ligands to two carbohydrate-binding proteins was quantified. Importantly, the carbohydrate affinities were found to be in good agreement with values measured using ITC. It is anticipated that this method will prove particularly useful in extending the application of ESI-MS binding measurements to protein interactions with small or hydrophobic ligands and for implementing a small fragment approach to elucidating the details of protein-oligosaccharide interactions.[23]

3.5 Literature cited

1. Daniel, J. M.; Friess, S. D.; Rajagopalan, S.; Wendt, S.; Zenobi, R. *Int. J. Mass Spectrom.* **2002**, *216*, 1-27.
2. Schug, K. A. *Comb. Chem. High Throughput Screen.* **2007**, *10*, 301-316.
3. Heck A. J. R.; Van Den Heuvel R. H. H. *Mass Spectrom. Rev*. **2004**, *23*, 368-389.
4. Wang, W.; Kitova, E. N.; Klassen, J. S. *Anal. Chem*. **2003**, *75*, 4945-4955.
5. Shoemaker, G. K.; Soya, N.; Palcic, M. M.; Klassen, J. S. *Glycobiology* **2008**, *18*, 587-592.
6. Soya, N.; Shoemaker, G. K.; Palcic, M. M.; Klassen, J. S. *Glycobiology* **2009**, *19*, 1224-1234.
7. Rademacher, C.; Shoemaker, G. K.; Kim, H. S.; Zheng, R. B.; Taha, H.; Liu, C.; Nacario, R. C.; Schriemer, D. C.; Klassen, J. S.; Peters T.; Lowary, T. L. *J. Am. Chem. Soc.* **2007**, *129*, 10489-10502.
8. Jorgensen, T. J. D.; Roepstorff, P.; Heck, A. J. R. *Anal*. *Chem*. **1998**, *70*, 4427-4432.
9. Jecklin, M. C.; Touboul, D.; Bovet, C.; Wortmann, A.; Zenobi, R. *J*. *Am*. *Soc*. *Mass Spectrom*. **2008**, *19*, 332-343.
10. Yu, Y. H.; Kirkup, C. E.; Pi, N.; Leary, J. A. *J*. *Am*. *Soc*. *Mass Spectrom*. **2004**, *15*, 1400-1407.
11. Van Dongen, W. D.; Heck A. J. R. *Analyst* **2000**, *125*, 583-589.
12. Sun, J.; Kitova, E. N.; Klassen, J. S. *Anal*. *Chem*. **2007**, *79*, 416-425.
13. Clark, S. M.; Konermann, L. *Anal*. *Chem*. **2004**, *76*, 7077-7083.

14. Robinson, C. V.; Chung, E. W.; Kragelund, B. B.; Knudsen, J.; Aplin, R. T.; Poulsen, F. M.; Dobson, C. M. *J. Am. Chem. Soc.* **1996**, *118*, 8646-8653.

15. Bagal, D.; Kitova, E. N.; Liu, L.; El-Hawiet, A.; Schnier, P. D.; Klassen, J. S. *Anal*. *Chem*. **2009**, *81*, 7801-7806.

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

17. 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*, 5172-5182.

18. Bundle, D. R. Unpublished results.

19. Castro, S.; Duff, M.; Snyder, N. L.; Morton, M.; Kumar, C. V.; Peezuh, M. W. *Org. Biomol. Chem*. **2005**, *3*, 3869-3872.

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

21. Mandal, D. K.; Brewer, C. F. *Biochemistry* **1993**, *32*, 5116-5120.

22. Loris, R.; Maes, D.; Poortmans, F.; Wyns, L.; Bouckaert, J. *J. Biol. Chem*. **1996**, *271*, 30614-30618.

23. De Kloe, G. E.; Bailey, D.; Leurs, R.; De Elch, I. J. P. *Drug Discovery Today* **2009**, *14*, 630-646.

Chapter 4

Quantifying Ligand Binding to Large Protein Complexes Using Electrospray Ionization Mass Spectrometry[*+]

4.1 Introduction

The direct electrospray ionization mass spectrometry (ESI-MS) assay is increasingly used to quantify protein-ligand interactions, as well as other noncovalent biological complexes, in aqueous solution.[1-13] The assay is based on the direct detection and quantification of the gas phase ions of ligand-bound and unbound protein, e.g., PL and P, respectively. An underlying assumption of this assay is that the ratio (*R*) of the abundance (*Ab*) of ligand-bound and free protein ions measured by ESI-MS is equivalent to the equilibrium concentration ratio in solution, eq 1.3. From the measured *R* value and initial concentrations of protein and ligand ($[P]_o$ and $[L]_o$, respectively), the association constant (K_a) can be calculated, eq 1.2.[4]

The requirement that the free and ligand-bound protein ions can be detected and accurately quantified places restrictions on the nature of the interactions that can be measured using the direct ESI-MS assay. Currently, affinity measurements involving relatively small ligands, with molecular weights (MW) of a few hundred Da, are generally restricted to proteins and protein complexes with MWs <200 kDa. And,

[*] A version of this chapter has been published: El-Hawiet, A.; Kitova, E. N.; Arutyunov, D.; Simpson, D. J.; Szymanski, C. M.; Klassen, J. S. *Anal*. *Chem*. **2012**, *84*, 3867–3870.

[+] Protein expression and purification were done by Arutyunov, D.and Simpson, D. J (University of Alberta)

depending on the mass analyzer used, the practical upper mass limit may be significantly lower. Protein microheterogeneity may also hinder the implementation of the direct ESI-MS assay. In cases where direct analysis is not feasible, it may be possible to quantify protein-ligand interactions using indirect MS methods.[13-22]

Here, we describe an approach, which combines the direct ESI-MS assay and competitive protein-ligand binding, to quantify ligand interactions with high MW proteins and protein complexes that are not readily directly detected by ESI-MS. The method, which we refer to as the *proxy protein* ESI-MS method, employs a proxy protein (P_{proxy}) that binds specifically to the ligand of interest with known affinity. The fraction of P_{proxy} bound to L, which is determined directly from the ESI mass spectrum, is sensitive to the fraction of P bound to L in solution and enables the affinity of the PL complex to be determined. A modified form of the *proxy protein* ESI-MS method, which accounts for real-time changes in ligand concentration, is also described. To validate the methods, the interactions of the wildtype (WT) and endorhamnosidase single point mutant (D392N) of the 180 kDa homotrimeric tailspike protein (TSP) of the bacteriophage P22[23-25] with two carbohydrate ligands,[26,27] an octasaccharide (**O** ≡ [α-Gal-(1→2)-[α-Abe-(1→3)]-α-Man-(1→4)-α-Rha]$_2$) and a dodecasaccharide (**D** ≡ [α-Gal-(1→2)-[α-Abe-(1→3)]-α-Man-(1→4)-α-Rha]$_3$) comprising two and three O-antigen repeats from *Salmonella typhimurium*, were quantified at 10 and 25 °C. The results were compared with K_a values measured previously using a fluorescence quenching assay.[23]

4.2 Experimental

4.2.1 Proteins and ligands

Truncated versions of the wildtype (WT, monomer MW 61386.9 Da) and endorhamnosidase inactivated (D392N, monomer MW 61385.9 Da)[23] P22 bacteriophage tailspike (TSP) genes lacking the codons for the first 108 amino acids comprising the phage head binding domain, were generated as previously described.[24,25] These wildtype and mutant forms of P22 TSP were expressed with a His-tag and purified using immobilized metal affinity chromatography. The proteins were diluted 10 fold into phosphate-buffered saline and precipitated with 30% ammonium sulphate. The TSPs were then re-suspended in 100 mM ammonium acetate and dialyzed (Thermo Scientific Slide-A-Lyzer cassettes, 3500 MWCO) against ammonium acetate at 4°C to remove any residual PBS prior to use. The carbohydrate-binding single chain variable fragment (scFv, MW 26 539 Da) of the monoclonal antibody Se155-4 was produced using recombinant technology.[28] Each protein was concentrated and dialyzed against aqueous 50 mM ammonium acetate and stored at 4 ^{o}C, if not used immediately. The octasaccharide (**O**, [α-Gal-(1→2)-[αAbe-(1→3)]-α-Man-(1→4)-α-Rha]$_2$) and dodecasaccharide (**D**, [α-Gal-(1→2)-[αAbe-(1→3)]-α-Man-(1→4)-α-Rha]$_3$ ligands (Figure 4.1) were generously provided by Prof. D. Bundle (University of Alberta). The oligosaccharides were produced by phage P22 hydrolysis of O-antigen polysaccharide chains of *Salmonella* serogroup B lipopolysaccharides followed by gel filtration.[26,27] Shown in Figure 4.2 are representative ESI mass spectra acquired for the aqueous solutions of **O** and **D**. According to ESI-MS analysis, the sample of **O** is predominantly the expected octasaccharide (Figure 4.2). However, the

sample of **D** contains an appreciable amount of unidecasaccharide (**U**) (Figure 4.1), a hydrolysis product from the terminal non-reducing end of the **O** chain.[27] Assuming similar ESI-MS response factors for **D** and **U**, the fractional abundance of **U** in the sample was estimated to be 0.35. Efforts to separate the oligosaccharides by size exclusion chromatography were unsuccessful. The affinities of the scFv for **D** and **U** were analyzed by direct ESI-MS. Based on their estimated concentrations, the mass spectral data indicate that **D** and **U** bind to scFv with similar affinities at pH 7 and 25 °C, $(2.1 \pm 0.3) \times 10^5$ and $(1.8 \pm 0.1) \times 10^5$, respectively. Therefore**,** to implement the *proxy protein* method, **D** and **U** were treated as a single species, with an effective concentration reflective of the fractional abundance of the two components and their individual molecular weights.

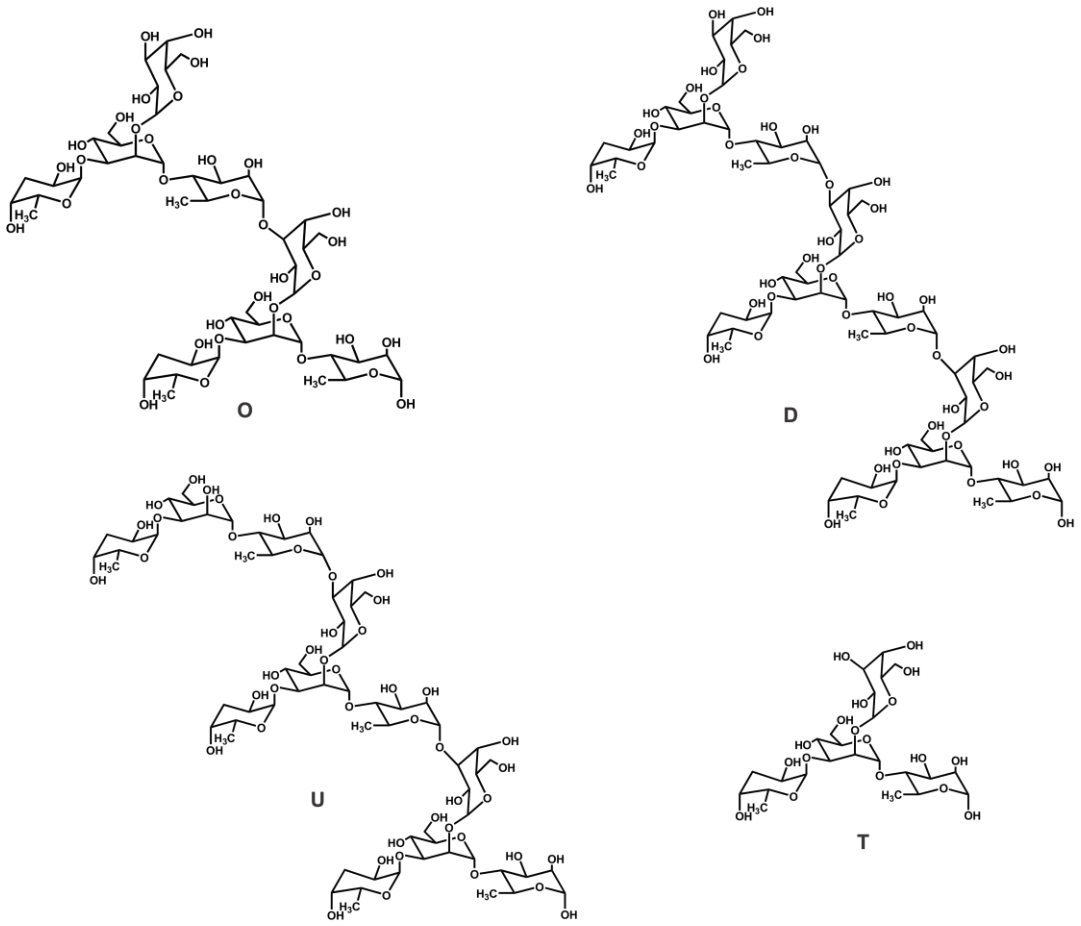


Figure 4.1 Structures of carbohydrate ligands: dodecasaccharide, **D**; unidecasaccharide, **U**; octasaccharide, **O**; and tetrasaccharide, **T**.

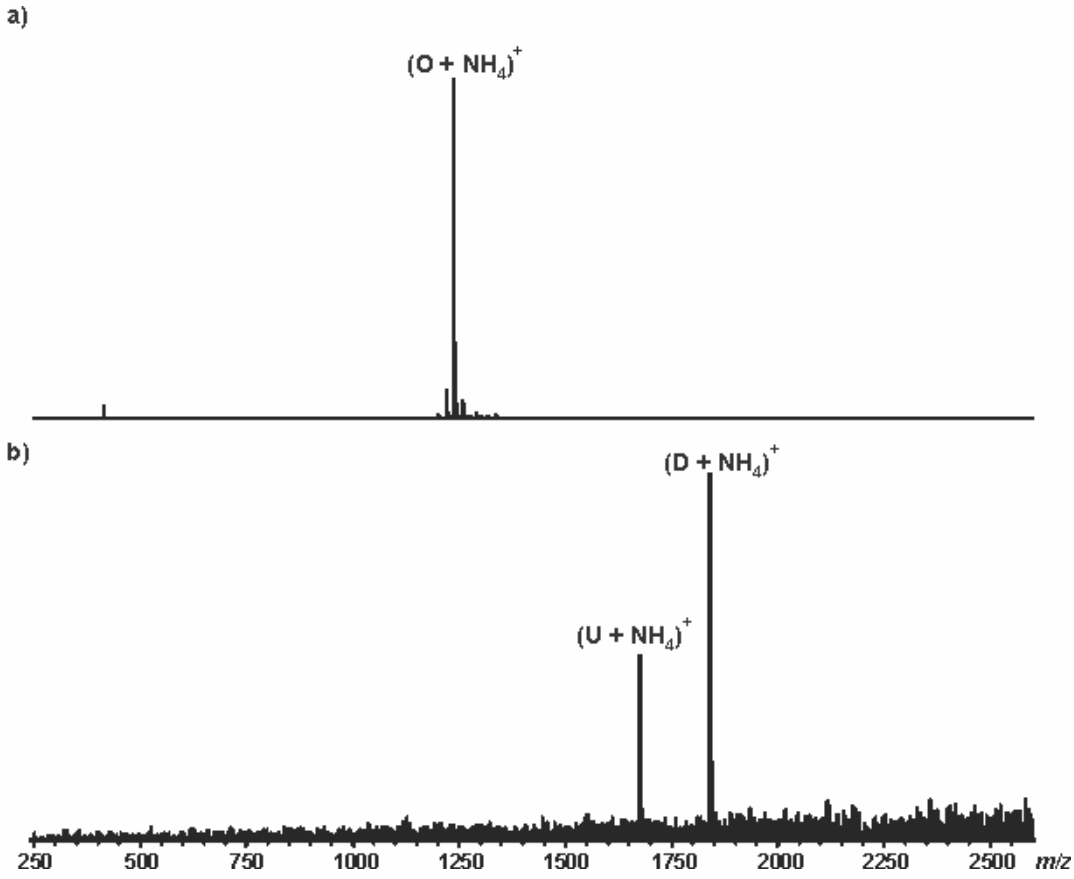


Figure 4.2 ESI mass spectra obtained for aqueous solutions of (a) **O** (10 μM) and (b) **D** (10 μM). According the ESI mass spectrum, the sample of **D** is found to contain approximately 35% **U**.

4.2.2 Mass spectrometry

The binding measurements at 10 °C were carried out with a 9.4T ApexII Fourier transform ion cyclotron resonance (FTICR) mass spectrometer (Bruker, Billerica, MA), while the measurements at 25 °C were performed using a 9.4T ApexQe FTICR mass spectrometer (Bruker, Billerica, MA). In both cases, nanoflow ESI (nanoESI) was performed using borosilicate tubes (1.0 mm o.d., 0.68 mm i.d.), pulled to ~5 μm o.d. at one end using a P-2000 micropipette puller (Sutter Instruments, Novato, CA). The 10 °C binding measurements were performed using a home-built temperature-controlled nanoESI device.[29]

ApexII 9.4T FTICR mass spectrometer. Details of the instrumental and experimental conditions used for the protein-ligand affinity measurements performed with this instrument are given elsewhere[4] and described in Chapter 1.

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 voltage was held at 0 V, and the capillary exit was 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 512K data points per transient were used for each acquisition.

4.3 Results and discussion

4.3.1 Overview of proxy protein ESI-MS method

The *proxy protein* ESI-MS method requires a suitable P_{proxy}, one that binds specifically to L with known affinity, and for which the interaction can be quantified using the direct ESI-MS assay. In the absence of P, the measured abundance ratios of ligand-bound to free P_{proxy} ions (i.e., R_{proxy}) will reflect the concentration ratio ($[P_{proxy}L]/[P_{proxy}]$) and K_a for the $P_{proxy}L$ interaction (i.e., $K_{a,proxy}$) according to eq 4.2. The addition of P to the solution will result in a reduction in the magnitude of $[P_{proxy}L]$ and R_{proxy} due to the formation of the PL complex. Although the value of $K_{a,P}$ (the microscopic association constant for the PL complex) cannot be determined directly from the ESI mass spectrum, it can be calculated from the measured value of R_{proxy} and the equation of mass balance, eq 4.1:

$$[L]_o = [L] + [P_{proxy}L] + \sum_{1 \le h \le i} h[PL_h] \tag{4.1}$$

where i is the total number of ligand binding sites in P and $\sum_{1 \le h \le i} h[PL_h]$ is the total concentration of L bound to P. The concentration term [L] can be found from eq 4.2:

$$[L] = \frac{R_{proxy}}{K_{a,proxy}} \tag{4.2}$$

and the concentration term $[P_{proxy}L]$ can be calculated from eq 4.3:

$$[P_{proxy}L] = \frac{[P_{proxy}]_o R_{proxy}}{R_{proxy} + 1} \tag{4.3}$$

Substituting the values of [L] and [$P_{proxy}L$] into eq 4.3 gives $\sum_{1\le h\le i} h[PL_h]$, as shown in eq 4.4:

$$\sum_{1\le h\le i} h[PL_h] = [L]_o - \frac{[P_{proxy}]_o R_{proxy}}{R_{proxy}+1} - \frac{R_{proxy}}{K_{a,proxy}} \tag{4.4}$$

Since the total concentration of ligand binding sites is $i[P]_o$ and the $\sum_{1\le h\le i} h[PL_h]$ term corresponds to the molar concentration of occupied binding sites, $K_{a,P}$ can be calculated using eq 4.5a:

$$K_{a,P} = \frac{\sum_{1\le h\le i} h[PL_h]}{(i[P]_o - \sum_{1\le h\le i} h[PL_h])[L]} \tag{4.5a}$$

which, after substituting in the expressions for [L] and $\sum_{1\le h\le i} h[PL_h]$, according to eqs 4.2 and 4.4, takes the following form, eq 4.5b:

$$K_{a,P} = \frac{K_{a,proxy}([L]_o - \frac{[P_{proxy}]_o R_{proxy}}{R_{proxy}+1} - \frac{R_{proxy}}{K_{a,proxy}})}{(i[P]_o - L_o + \frac{[P_{proxy}]_o R_{proxy}}{R_{proxy}+1} + \frac{R_{proxy}}{K_{a,proxy}})R_{proxy}} \tag{4.5b}$$

If *i* is known, $K_{a,P}$, can, in principle, be determined from a single ESI mass spectrum. Alternatively, if *i* is not known, measurements must be performed at two or more concentrations in order to simultaneously determine *i* and $K_{a,P}$. In practice, it is convenient to use a titration approach, in which measurements are carried out on solutions with fixed values of $[L]_o$ and $[P_{proxy}]_o$ and varying $[P]_o$ values.

To demonstrate the reliability of the *proxy protein* ESI-MS method, the assay was used to quantify the binding of P22 TSP WT to **O** and D392N to **O** and **D** and the

results were compared with those obtained using a fluorescence quenching assay.[23] The single chain variable fragment (scFv) of the monoclonal antibody Se155-4[28] served as P_{proxy} for these measurements. Using the direct ESI-MS assay, the scFv was found to bind to **O** and **D** at 10 °C, pH 7 with affinities of (1.3 ± 0.1) x 10^6 M^{-1} and (1.3 ± 0.3) x 10^6 M^{-1}, respectively and at 25 °C, pH 7 with affinities of (4.3 ± 1.0) x 10^5 M^{-1} and (5.1 ± 0.6) x 10^5 M^{-1}, respectively. Notably, the values measured at 25 °C are in good agreement with values determined by isothermal titration calorimetry (**O**, (5.3 ± 1.6) x 10^5 M^{-1}; **D** (4.9 ± 3.0) x 10^5 M^{-1})).[26] Shown in Figure 4.3a is a representative ESI mass spectrum acquired for a solution of scFv (5 μM) and **O** (5 μM) at 25 °C. Both free ($scFv^{n+}$) and ligand-bound scFv ($(scFv + \mathbf{O})^{n+}$) ions, at n = 8 - 11, were detected. At these concentrations, 54% of the scFv is bound to **O**; this translates to a R_{proxy} value of 1.17. Shown in Figures 4.3b and 4.3c are ESI mass spectra obtained for the same solution, but with the addition of 1.8 μM and 5.4 μM of P22 TSP WT, respectively. The addition of WT to the solution results in a measurable decrease in the fraction of bound scFv, to 51% (R_{proxy} = 1.04) and 47% (R_{proxy} = 0.89). Due to the high MW of WT, no ions corresponding to free or ligand-bound WT could be detected with the mass spectrometer used for these measurements. Nevertheless, the decrease in the fraction of ligand-bound scFv confirms that WT is competing with scFv for **O** in solution. In order to establish the K_a for WT binding to **O**, a titration experiment was performed, wherein the concentrations of scFv (5 μM) and **O** (5 μM) were fixed and the concentration of WT was varied (from 0 to 7 μM). From the measured R_{proxy} values and taking into account the three ligand binding sites (i.e., i = 3) of the TSP trimer, [23-25]

an average K_a of (2.9 ± 0.9) x 10^5 M^{-1} was obtained (Table 4.1). Notably, this value agrees within a factor of two of the reported value, 6.0 x 10^5 M^{-1}.[23]

Table 4.1 Association constants (K_a) for carbohydrate (**O** and **D**) binding to P22 TSP wildtype (WT) and mutant (D392N) determined at 10 and 25°C and pH 7 by the *proxy protein* ESI-MS method and by a fluorescence quenching (FQ) assay.[a,b]

TSP	Ligand	ESI-MS 10 °C K_a x 10^5 (M^{-1})	ESI-MS 25 °C K_a x 10^5 (M^{-1})	FQ 10 °C [b] K_a x 10^5 (M^{-1})	FQ 25 °C [b] K_a x 10^5 (M^{-1})
WT	**O**	10.2 ± 2.0	2.9 ± 0.9	11.5 ± 3.6	6.0
WT	**D**	-	4.5 ± 1.9	-	-
D392N	**O**	10.9 ± 3.0	3.2 ± 0.7	9.5 ± 1.7	4.0
D392N	**D**	24.3 ± 5.1	5.0 ± 1.7	23.8 ± 2.8	6.9

a. Errors correspond to one standard deviation. b. Values from reference 23.

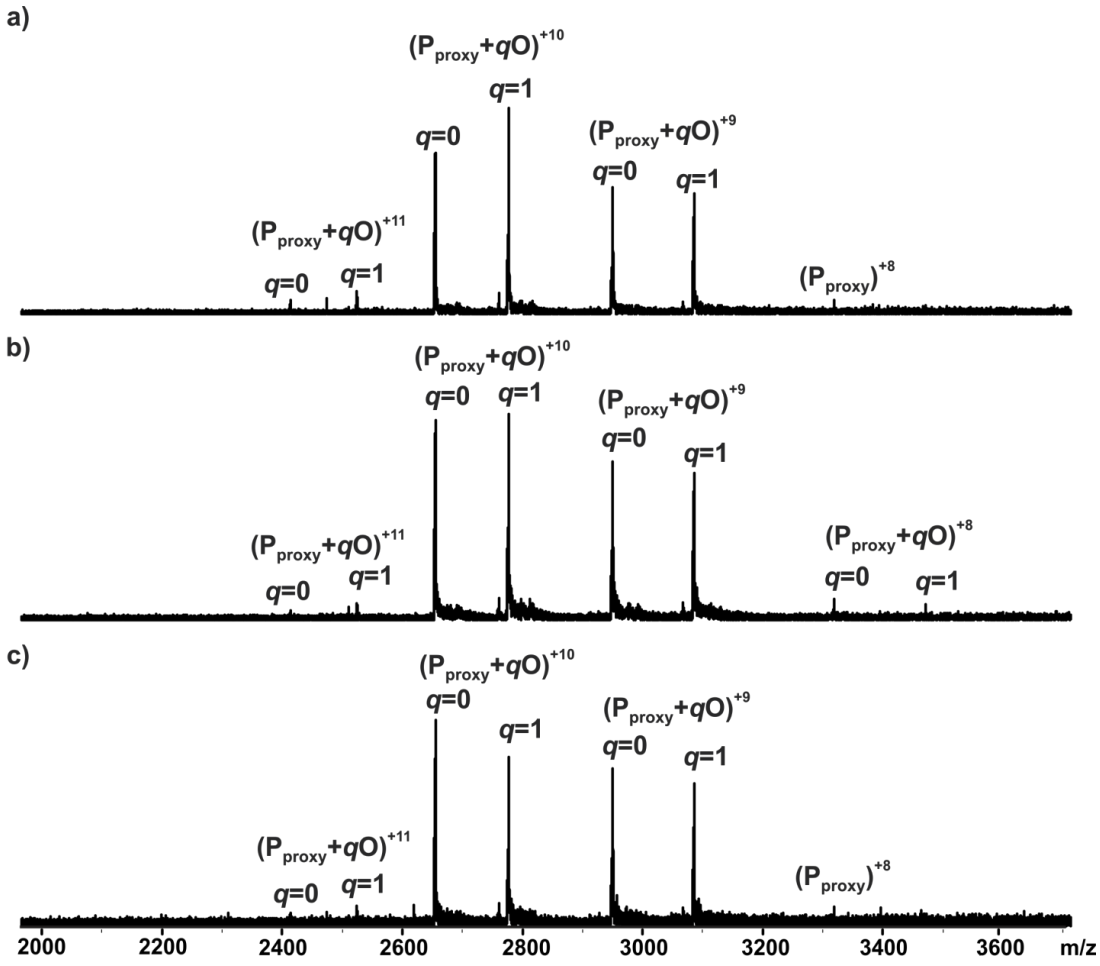


Figure 4.3 ESI mass spectra obtained for aqueous solutions (25 ^{o}C at pH 7) of scFv ≡ P_{proxy} (5.0 μM), **O** (5.0 μM) and P22 TSP WT at (a) 0.0 μM (b) 1.8 μM and (c) 5.4 μM. The number of molecules of **O** bound to the P_{proxy} ions is indicated by *q*.

Following the same procedure, the affinities of P22 TSP D392N for **O** and **D** were determined at 25 ^{o}C (Figures 4.4 and 4.5). The corresponding plots of R_{proxy} versus concentration are also shown in Figure 4.6a. Analysis of the experimental data yields K_a values of (5.0 ± 1.7) x 10^5 M^{-1} and (3.2 ± 0.7) x 10^5 M^{-1}, for **O** and **D**, respectively (Table 4.1). Again, these values are in good agreement with the reported values for **D** (6.9 x 10^5 M^{-1}) and **O** (4.0 x 10^5 M^{-1}).[23] Using a temperature-controlled ESI device,[29] titration experiments were also performed at 10 °C to establish affinities for WT with **O** and D392N with **O** and **D** (Figures 4.7-4.9). The resulting plots of R_{proxy} versus concentration are shown in Figure 4.6b and the corresponding K_a values are listed in Table 4.1. For all three interactions, the K_a values determined with the *proxy protein* method ((10.2 ± 2.0) x 10^5 M^{-1} for WT with **O** and (10.9 ± 3.0) x 10^5 M^{-1} and (24.3 ± 5.1) x 10^5 M^{-1} for D392N with **O** and **D**), respectively, agree with values measured using the fluorescence quenching assay ((11.5 ± 3.6) x 10^5 M^{-1} for WT with **O** and (9.5 ± 1.7) x 10^5 M^{-1} and (23.8 ± 2.8) x 10^5 M^{-1} for D392N with **O** and **D** , respectively).[23]

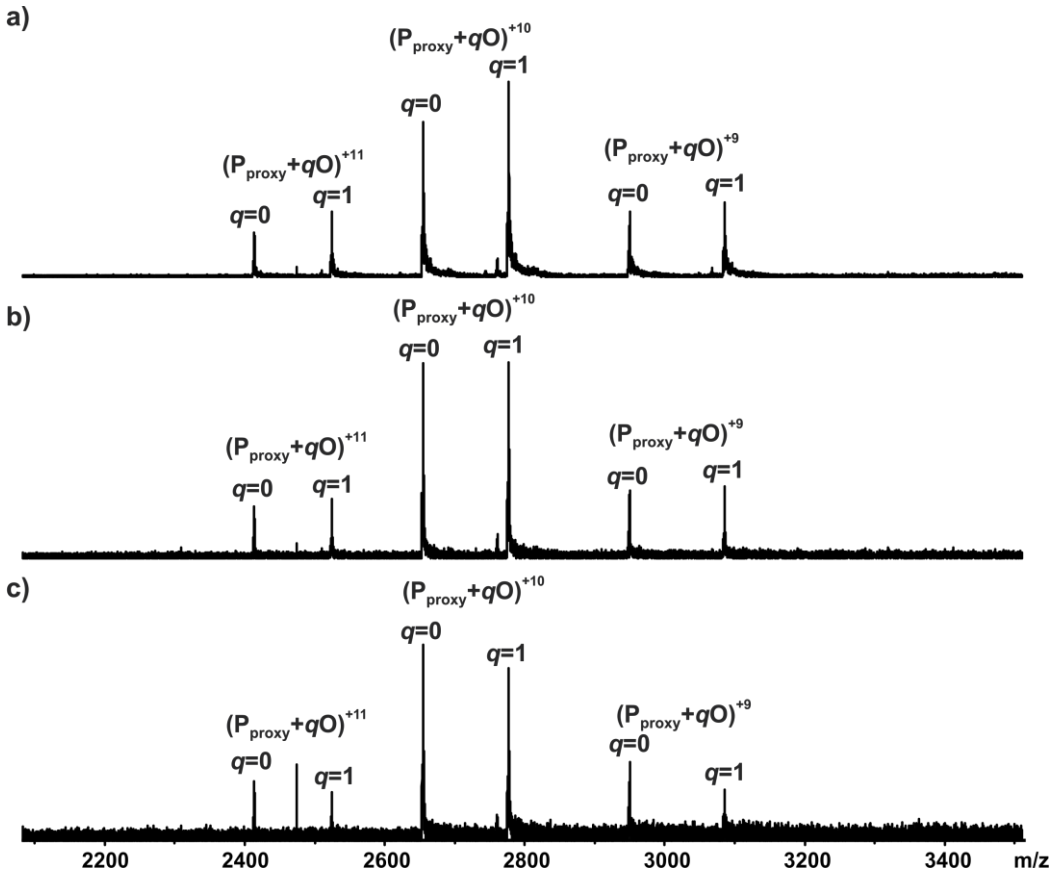


Figure 4.4 ESI mass spectra obtained, at 25 ^{o}C, for the solutions of Se155-4 scFv (5.0 μM), **O** (5.0 μM) and P22 TSP D392N at (a) 0.0 μM (b) 1.2 μM and (c) 1.5 μM. The number of molecules of **O** bound to the protein ions is indicated by *q*.

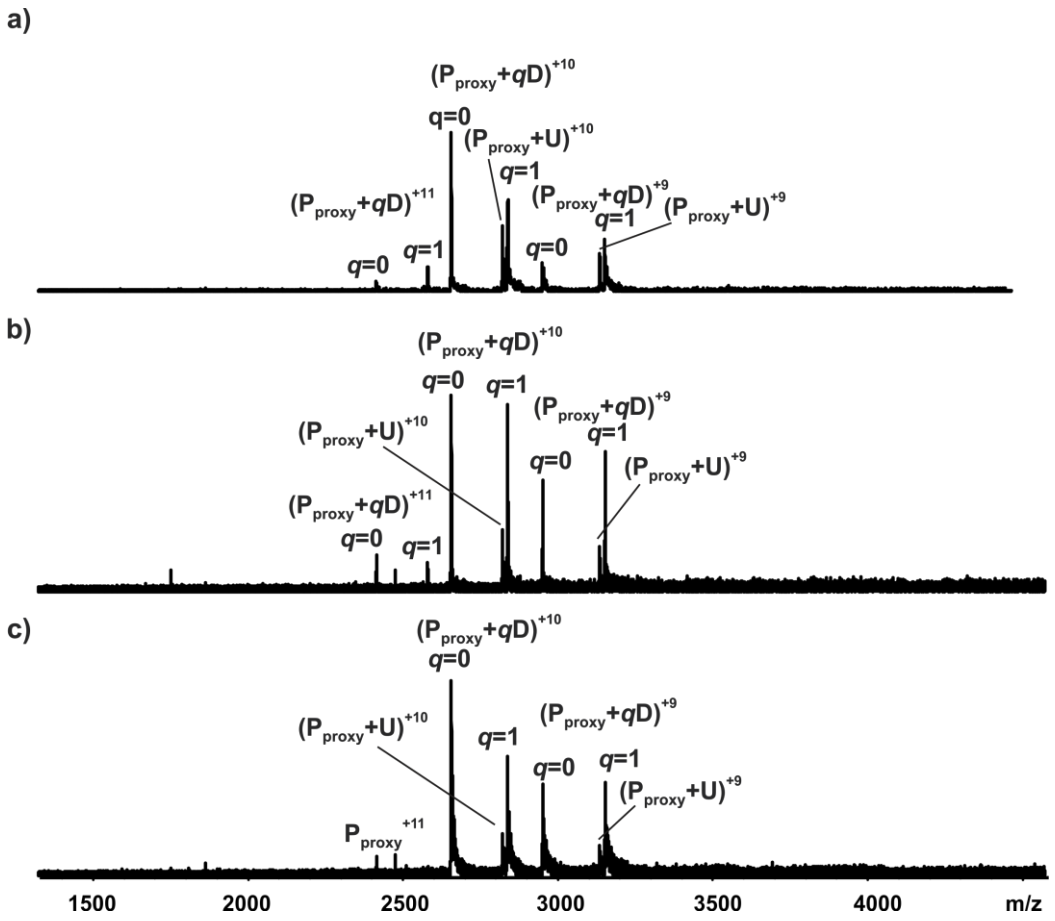


Figure 4.5 ESI mass spectra obtained, at 25 ^{o}C, for the solutions of Se155-4 scFv (1.5 μM), **D** (5.0 μM) and P22 TSP D392N at (a) 0.0 μM (b) 0.9 μM and (c) 1.9 μM. The number of molecules of **D** bound to the protein ions is indicated by *q*.

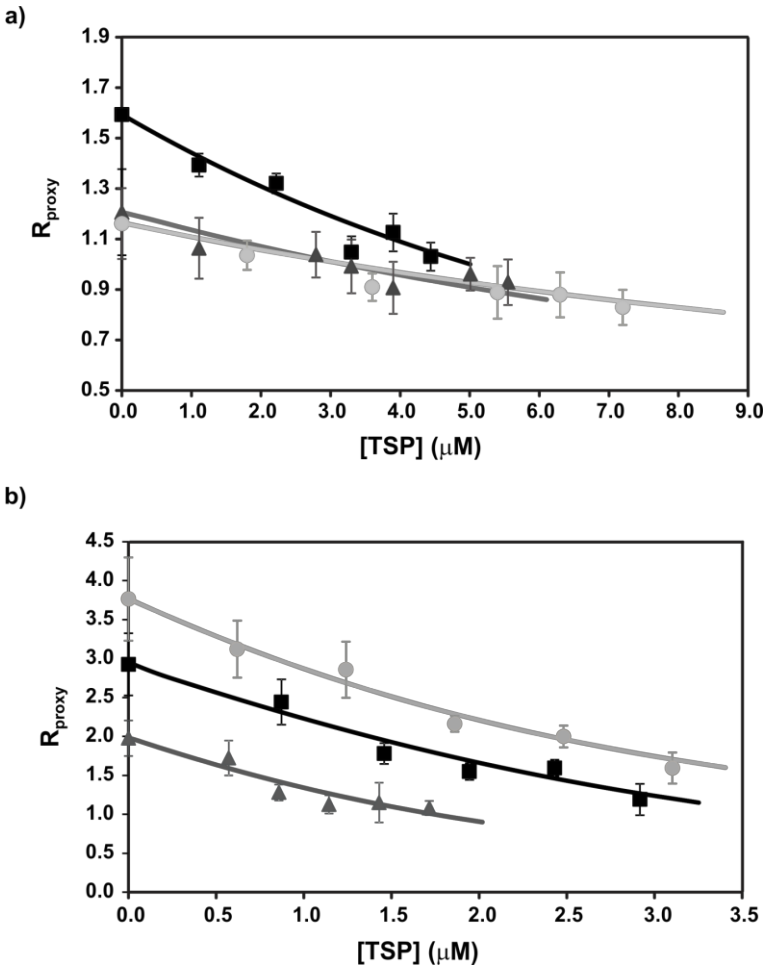


Figure 4.6 Plots of R_{proxy} versus P22 TSP concentrations measured at (a) 25 ^{o}C at pH 7 for a solution of scFv ≡ P_{proxy} (5.0 μM), carbohydrate ligand (5.0 μM) and TSP (0.0 – 7.2 μM): ●, WT and **O**; ■, D392N and **D**; ▲, D392N and **O**, and at (b) 10 ^{o}C at pH 7 for a solution of scFv ≡ P_{proxy} (5.0 μM), carbohydrate ligand (5.0 μM with WT or 3.5 μM with D392N) and TSP (0.0 – 3.3 μM): ●, WT and **O**; ■, D392N and **D**; ▲, D392N and **O**. The solid curves describe the concentration dependence of R_{proxy} expected based on the average K_a values (Table 4.1) determined by ESI-MS for each interaction.

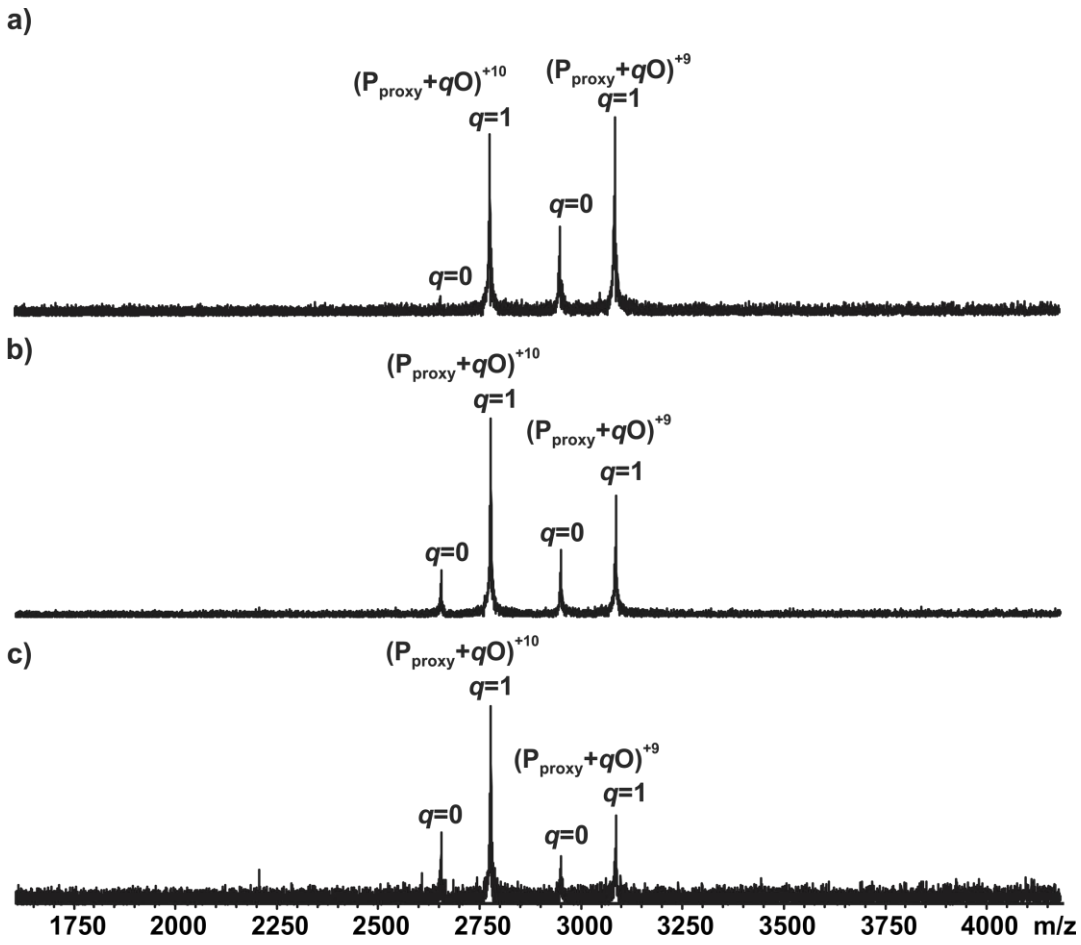


Figure 4.7 ESI mass spectra obtained, at 10°C, for the solutions of Se155-4 scFv (5.0 μM), **O** (5.0 μM) and P22 TSP WT at (a) 0.0 μM (b) 1.2 μM and (c) 1.5 μM. The number of molecules of **O** bound to the protein ions is indicated by *q*.

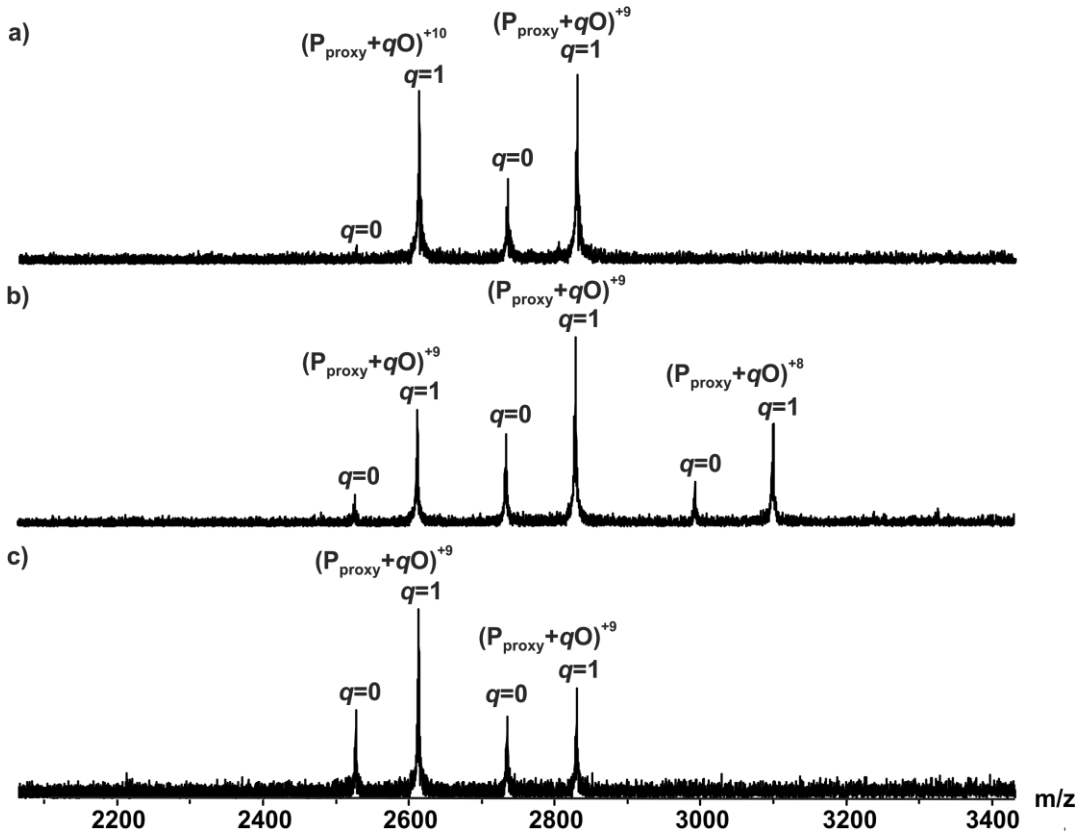


Figure 4.8 ESI mass spectra obtained, at 10°C, for the solutions of Se155-4 scFv (3.0 μM), **O** (5.0 μM) and P22 TSP D392N at (a) 0.0 μM (b) 1.1 μM and (c) 1.7 μM. The number of molecules of **O** bound to the protein ions is indicated by *q*.

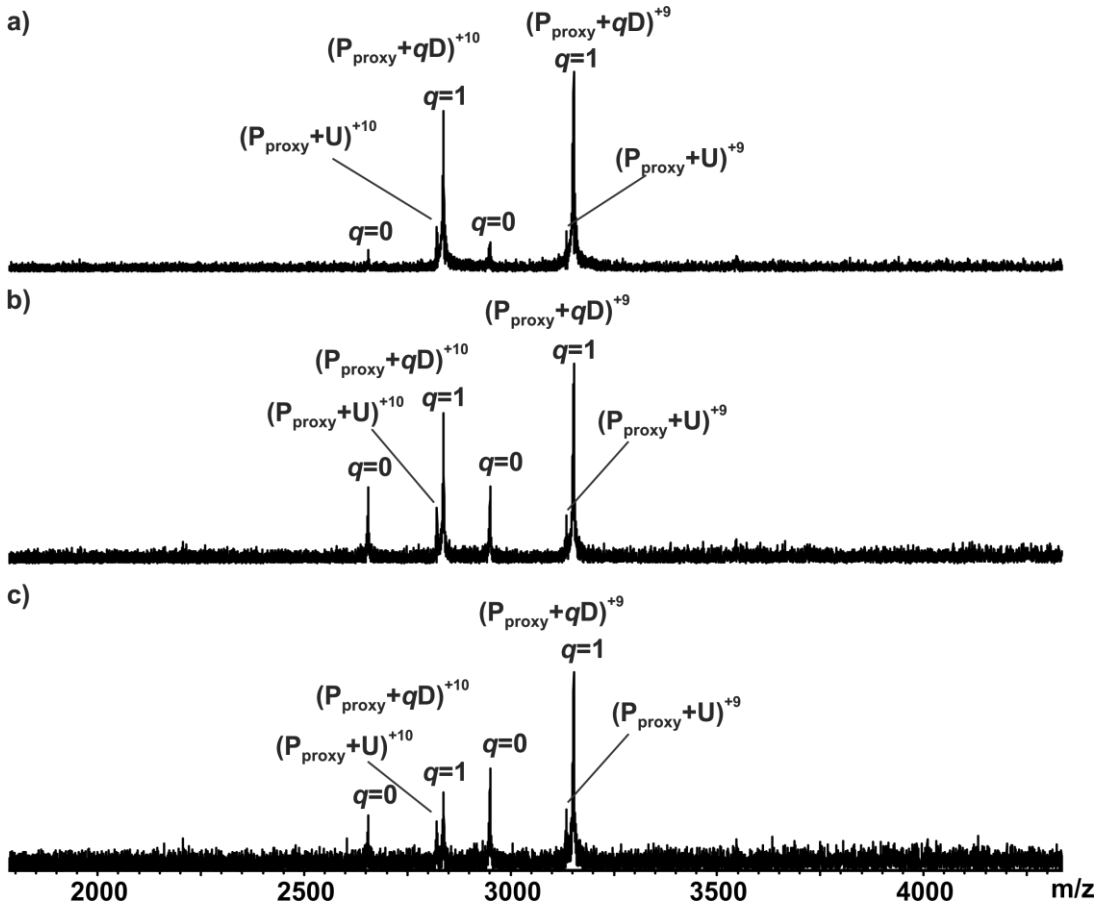


Figure 4.9 ESI mass spectra obtained, at 10°C, for the solutions of Se155-4 scFv (1.5 μM), **D** (5.0 μM) and P22 TSP D392N at (a) 0.0 μM (b) 0.9 μM and (c) 1.9 μM. The number of molecules of **D** bound to the protein ions is indicated by *q*.

4.3.2 Application of the proxy protein ESI-MS method to enzyme-substrate interactions

The P22 TSP WT possess endorhamnosidase hydrolyzing activity towards **D**, which results in the formation of **O** and the corresponding tetrasaccharide (**T**, α-Gal-(1→2)-[α-Abe-(1→3)]-α-Man-(1→4)-α-Rha). Consequently, a solution of WT and **D** will also contain **O**, which binds to WT, and **T**, which does not bind to WT. As a result, the *proxy protein* ESI-MS method cannot be used in the manner described above to evaluate the affinity of WT for **D**. However, a special application of the *proxy protein* ESI-MS method, which was utilized in the present study, involves quantifying PL interactions for which the concentration of L is not constant. The specific case considered here involves an interaction between an enzyme (P) and its substrate (L1), which is converted to L2 and L3, eq 4.6:

$$P + L1 \rightarrow P + L2 + L3 \tag{4.6}$$

L2 and L3 are not substrates but can bind noncovalently to P (and P_{proxy}). For a P with *i* binding sites for L1, L2 and L3, a P_{proxy} with a single binding site for L1, L2 and L3, and assuming only L1 is present initially (i.e., $[L2]_o = [L3]_o = 0$), the relevant mass balance equations are:

$$i[P]_o = i[P] + \sum_{1 \le h \le i} h[PL1_h] + \sum_{1 \le h \le i} h[PL2_h] + \sum_{1 \le h \le i} h[PL3_h] \tag{4.7}$$

$$[P_{proxy}]_o = [P_{proxy}] + [P_{proxy}L1] + [P_{proxy}L2] + [P_{proxy}L3] \tag{4.8}$$

$$[\mathrm{L1}]_\mathrm{o} = [\mathrm{L1}] + [\mathrm{L2}] + [\mathrm{L3}] + \sum_{1\le h\le i} h[\mathrm{PL1}_h] + [\mathrm{P_{proxy}L1}] + \sum_{1\le h\le i} h[\mathrm{PL2}_h] + [\mathrm{P_{proxy}L2}] + \sum_{1\le h\le i} h[\mathrm{PL3}_h] + [\mathrm{P_{proxy}L3}] \quad (4.9)$$

The concentrations of free L (L1, L2 or L3) at equilibrium can be calculated from the corresponding association constant for P_{proxy} and the measured abundance ratio of ligand-bound and free P_{proxy} ions, eqs 4.10a-c:

$$[\mathrm{L1}] = \frac{R_{\mathrm{proxy,L1}}}{\mathrm{K}_{\mathrm{a,proxy,L1}}} \quad (4.10a)$$

$$[\mathrm{L2}] = \frac{R_{\mathrm{proxy,L2}}}{\mathrm{K}_{\mathrm{a,proxy,L2}}} \quad (4.10b)$$

$$[\mathrm{L3}] = \frac{R_{\mathrm{proxy,L3}}}{\mathrm{K}_{\mathrm{a,proxy,L3}}} \quad (4.10c)$$

while the concentrations of ligand-bound P_{proxy} can be determined from eqs 4.11a-c:

$$[\mathrm{P_{proxy}L1}] = \frac{[\mathrm{P_{proxy}}]_\mathrm{o} R_{\mathrm{proxy,L1}}}{1 + R_{\mathrm{proxy,L1}} + R_{\mathrm{proxy,L2}} + R_{\mathrm{proxy,L3}}} \quad (4.11a)$$

$$[\mathrm{P_{proxy}L2}] = \frac{[\mathrm{P_{proxy}}]_\mathrm{o} R_{\mathrm{proxy,L2}}}{1 + R_{\mathrm{proxy,L1}} + R_{\mathrm{proxy,L2}} + R_{\mathrm{proxy,L3}}} \quad (4.11b)$$

$$[\mathrm{P_{proxy}L3}] = \frac{[\mathrm{P_{proxy}}]_\mathrm{o} R_{\mathrm{proxy,L3}}}{1 + R_{\mathrm{proxy,L1}} + R_{\mathrm{proxy,L2}} + R_{\mathrm{proxy,L3}}} \quad (4.11c)$$

The concentration of free ligand binding sites in P (i[P]) can be found by rearranging eqs 4.7 and 4.9 and using the concentrations of free ligand and ligand-bound P_{proxy}:

$$i[\mathrm{P}] = i[\mathrm{P}]_o - [\mathrm{L1}]_o + [\mathrm{L1}] + [\mathrm{L2}] + [\mathrm{L3}] + [\mathrm{P_{proxy}L1}] + [\mathrm{P_{proxy}L2}] + [\mathrm{P_{proxy}L3}] \quad (4.13)$$

In order to establish the total concentration of L1 bound to P (i.e., $\sum_{1\le h\le i} h[\mathrm{PL1}_h]$), the concentrations of the complexes of P with L2 and L3 must be known. These can be determined from the corresponding K_a values (i.e., $K_{a,P,L2}$ and $K_{a,P,L3}$). In the present study, L1 corresponds to **D** and L2 and L3 correspond to **O** and **T** (Figure 4.1). The microscopic K_a for **O** (L2) binding to WT (P) was determined in a separate experiment; **T** (L3) does not to bind to WT. It follows for this case that $\sum_{1\le h\le i} h[\mathrm{PL3}_h]$ is equal to zero and eq 4.8 reduces to eq 4.13:

$$i[\mathrm{P}]_o = i[\mathrm{P}] + \sum_{1\le h\le i} h[\mathrm{PL1}_h] + \sum_{1\le h\le i} h[\mathrm{PL2}_h] \quad (4.13)$$

and $\sum_{1\le h\le i} h[\mathrm{PL2}_h]$ can be calculated using eq 4.14:

$$\sum_{1\le h\le i} h[\mathrm{PL2}_h] = i[\mathrm{P}][\mathrm{L2}]K_{a,P,L2} \quad (4.14)$$

The total concentration of L1 bound to P, $\sum_{1\le h\le i} h[\mathrm{PL1}_h]$, can be found using eq 4.15:

$$\sum_{1\le h\le i} h[\mathrm{PL1}_h] = i[\mathrm{P}]_o - i[\mathrm{P}] - iK_{a,P,L2}[\mathrm{P}][\mathrm{L2}] \quad (4.15)$$

Finally the microscopic K_a for P binding to L1, $K_{a,P,L1}$, is given by eq 4.16:

$$K_{a,P,L1} = \frac{\sum_{1\le h\le i} h[\mathrm{PL1}_h]}{i[\mathrm{P}][\mathrm{L1}]} = \frac{i([\mathrm{P}]_o - [\mathrm{P}] - K_{a,P,L2}[\mathrm{P}][\mathrm{L2}])K_{a,proxy,L1}}{R_{proxy,L1}(i[\mathrm{P}]_o - [\mathrm{L1}]_o + [\mathrm{L1}] + [\mathrm{L2}] + [\mathrm{L3}] + [\mathrm{P_{proxy}L1}] + [\mathrm{P_{proxy}L2}] + [\mathrm{P_{proxy}L3}])} \quad (4.16)$$

Using the modified form of the assay, it is possible to establish K_a for WT binding to **D** which is based on the fraction of P_{proxy} bound to **D**, **O** and **T** and the known affinity of WT for **O**. Shown in Figures 4.10a and 4.10b are representative ESI mass spectra acquired for a solution (25 ^oC at pH 7) of **D** (10 μM) and scFv (5 μM) in the absence of WT and approximately 5 min after the addition of WT (2.5 μM) to the solution, respectively. Prior to the addition of WT, only ion signal corresponding to free scFv and scFv bound to **D** and the unidecassacharide, **U,** i.e., $scFv^{n+}$, $(scFv + \mathbf{D})^{n+}$ and $(scFv + \mathbf{U})^{n+}$, at n = 9 and 10, is evident in the mass spectrum. As noted in experimental section, the sample of **D** contains (unavoidably)[27] a significant amount of **U** (Figure 4.2). After addition of WT, signal corresponding to $(scFv + \mathbf{O})^{n+}$ and $(scFv + \mathbf{T})^{n+}$ ions is also evident. In order to establish the K_a for WT binding to **D**, a titration experiment was performed in which the concentrations of scFv (5 μM) and **D** (10 μM) were fixed while the concentration of WT was varied. Analysis of the results yields an average K_a value of (4.5 ± 1.9) x 10^5 M^{-1}. Notably, this value, which is the first reported K_a value for P22 TSP WT binding to **D**, is identical, within experimental error, to the K_a value determined for D392N binding to **D**.

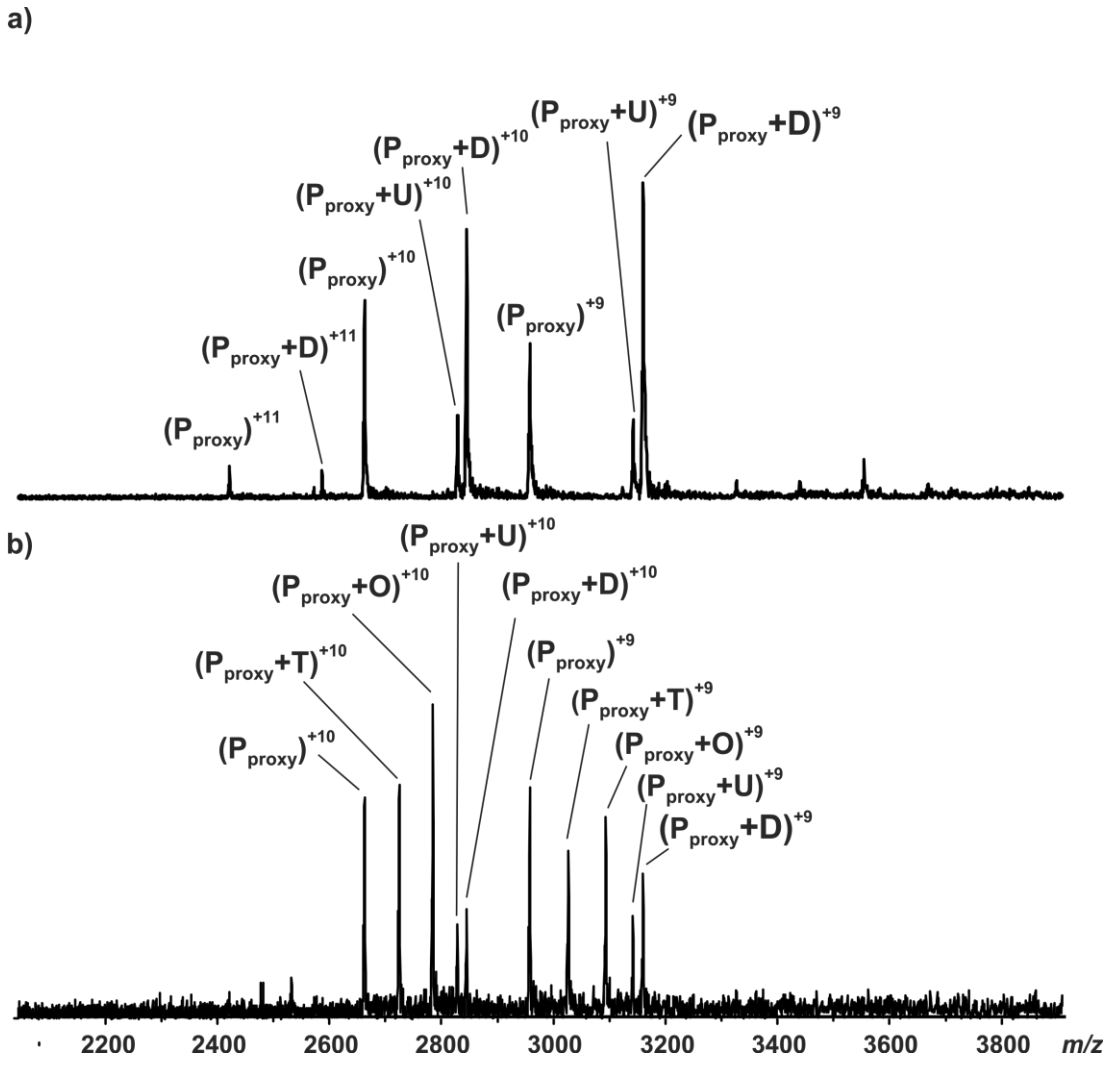


Figure 4.10 ESI mass spectra obtained for aqueous solutions (25 ^{o}C at pH 7) of scFv ≡ P_{proxy} (5.0 μM), **D** (10.0 μM) and P22 TSP WT at (a) 0 μM and (b) 2.5 μM. The carbohydrates **T** and **O** are products of the enzyme reaction involving WT and **D**. As described in experimental section, the sample of **D** used in this study contains approximately 35% unidecasaccharide, **U**.

4.4 Conclusions

In summary, a new ESI-MS binding assay, called the *proxy protein* ESI-MS method, for quantifying protein-ligand complexes that cannot be detected directly by ESI-MS, has been developed. A modified version of the assay, which accounts for real-time changes in ligand concentration, is also described. The reliability of the *proxy protein* ESI-MS method was demonstrated for the interactions between a 180 kDa homotrimeric tailspike protein of the bacteriophage P22, as well as an endorhamnosidase point mutant, and its octa- and dodecasaccharide ligands. Binding measurements performed using a single chain antibody as P_{proxy} at 10 and 25 °C yielded results that agree with reported K_a values. It should be noted that the *proxy protein* ESI-MS method is expected to be general and applicable to any protein system, regardless of size, provided a suitable P_{proxy} is available. It is anticipated that this method will prove particularly useful for the analysis of ligand interactions with very large protein assemblies, such as virus particles binding to their corresponding host-cell receptors, which are difficult to quantify using conventional assays.

4.5 Literature cited

1. Daniel, J. M.; Friess, S. D.; Rajagopalan, S.; Wendt, S.; Zenobi, R. *Int. J. Mass Spectrom.* **2002**, *216*, 1-27.
2. Schug, K. A. *Comb. Chem. High Throughput Screen.* **2007**, *10*, 301-316.
3. Heck A. J. R.; Van Den Heuvel R. H. H. *Mass Spectrom. Rev*. **2004**, *23*, 368-389.
4. Wang, W.; Kitova, E. N.; Klassen, J. S. *Anal. Chem*. **2003**, *75*, 4945-4955.
5. Shoemaker, G. K.; Soya, N.; Palcic, M. M.; Klassen, J. S. *Glycobiology* **2008**, *18*, 587-592.
6. Soya, N.; Shoemaker, G. K.; Palcic, M. M.; Klassen, J. S. *Glycobiology* **2009**, *19*, 1224-1234.
7. Rademacher, C.; Shoemaker, G. K.; Kim, H. S.; Zheng, R. B.; Taha, H.; Liu, C.; Nacario, R. C.; Schriemer, D. C.; Klassen, J. S.; Peters T.; Lowary, T.L. *J. Am. Chem. Soc.* **2007**, *129*, 10489-10502.
8. Jorgensen, T. J. D.; Roepstorff, P.; Heck, A. J. R. *Anal*. *Chem*. **1998**, *70*, 4427-4432.
9. Jecklin, M. C.; Touboul, D.; Bovet, C.; Wortmann, A.; Zenobi, R. *J*. *Am*. *Soc*. *Mass Spectrom*. **2008**, *19*, 332-343.
10. Yu, Y. H.; Kirkup, C. E.; Pi, N.; Leary, J. A. *J*. *Am*. *Soc*. *Mass Spectrom*. **2004**, *15*, 1400-1407.
11. Van Dongen, W. D.; Heck A. J. R. *Analyst* **2000**, *125*, 583-589.
12. Hofstadler, S. A.; Sannes-Lowery, K. A. *Nat*. *Rev*. *Drug Discov*. **2006,** 5,585-595.

13. Yu, Y.; Sweeney, M. D.; Saad, O. M.; Crown, S. E.; Hsu, A. R.; Handel, T. M.; Leary, J. A. *J*. *Biol*. *Chem*. **2005**, 280, 32200-32208.

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

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

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

17. Villanueva, J.; Yanes, O.; Querol, E.; Serrano, L.; Aviles, F. X. *Anal*. *Chem*. **2003**, 75, 3385–3395.

18. Yanes, O.; Aviles, F. X.; Roepstorff, P.; Jorgensen, T. J. *J*. *Am*. *Soc*. *Mass Spectrom*. **2006**, 18, 359-367.

19. Yanes, O.; Villanueva, J.; Querol, E.; Aviles, F. X. *Mol*. *Cell Proteomics* **2005**, 4, 1602–1613.

20. Wortmann, A.; Jecklin, M. C.; Touboul, D.; Badertscher, M.; Zenobi, R. *J*. *Mass Spectrom*. **2008**, *43*, 600-608.

21. Jecklin, M. C.; Touboul, D.; Jain, R.; Toole, E. N.; Tallarico, J.; Drueckes, P.; Ramage, P.; Zenobi, R. *Anal*. *Chem*. **2009**, *81*, 408–419.

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

23. Baxa, U.; Steinbacher, S.; Miller, S.; Weintraub, A.; Huber, R.; Seckler, R. *Biophys J.* **1996**, *71*, 2040–2048.

24. Henry, M. J.; Mackenzie, R. C.; Szymanski, C.; Jamshid, T. Phage receptor binding proteins for antibacterial therapy and other novel uses. PCT/US08/58675.

25. Waseh, S.; Hanifi-Moghaddam, P.; Coleman, R.; Masotti, M.; Ryan, S.; Foss, M.; MacKenzie, R.; Henry, M.; Szymanski, C. M.; Tanha, J. *PLoS ONE* **2010**, *5*, 1-9.

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

27. Eriksson, U.; Svenson, S. B.; Lonngren, J.; Lindberg, A. A. *J. Gen. Virol.* **1979**, *43*, 503-511.

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

29. Daneshfar, R.; Kitova, E. N.; Klassen, J. S. *J. Am. Chem. Soc.* **2004**, *126*, 4786-4787.

Chapter 5

Applications of a Catch and Release Electrospray Ionization Mass Spectrometry Assay for Carbohydrate Library Screening[*+]

5.1 Introduction

Carbohydrate-protein interactions play essential roles in numerous physiological and pathological processes. For example, such interactions are implicated in cell growth and differentiation, fertilization, in recognition processes, such as cell-cell adhesion and immune responses against pathogens, and in diverse disease mechanisms, including inflammatory processes and bacterial and viral adherence.[1-3] Understanding the molecular basis of carbohydrate-protein recognition, the relationship between structure and binding selectivity and affinity, is both of fundamental importance and facilitates the design of novel sugar-based therapeutic agents that may be used to treat a variety of diseases and infections.[4,5] Consequently, the development of analytical methods capable of identifying and quantifying biologically or therapeutically relevant protein-carbohydrate interactions represents an important and active area of research.

A variety of established analytical techniques exist for the discovery of protein-carbohydrate interactions in vitro, including surface plasmon resonance spectroscopy,[6] isothermal calorimetry,[7] frontal affinity chromatography mass spectrometry[8] and

[*] A version of this chapter has been published: El-Hawiet, A.; Shoemaker, G. K.; Daneshfar, R.; Kitova, E. N.; Klassen, J. S. *Anal*. *Chem*., **2012**, *84*, 50-58.

[+] Protein expression and purification were done by Blake Zheng (University of Alberta)

enzyme-linked immunosorbent assay.[9] Recently, a new technique, the glycan (carbohydrate) microarray has been adopted for screening carbohydrate libraries against target proteins or protein complexes.[10,11] The arrays consist of oligosaccharides immobilized on a solid support, usually through a covalent linkage. Carbohydrate-binding proteins are presented to the array and, following a washing step, the specific protein interactions are identified, usually using fluorescence detection. While the development of glycan microarrays has greatly improved current understanding of carbohydrate-protein recognition, there remain a number of challenges associated with the implementation of this technology. Among these are difficulties in controlling the glycan density in an immobilized state, the influence of the covalent linkers on binding specificity, and the lack of mobility of the glycans on the surface when using covalent linkers.[10,11]

Recently, the direct electrospray ionization (ESI) MS assay has emerged as a promising technique for evaluating the binding stoichiometries and affinities of protein-carbohydrate interactions, as well as other protein-ligand complexes, in vitro.[12-16] The assay is based on the direct detection and quantification of the relative abundance of free and ligand-bound protein ions measured for a solution of known initial protein and ligand concentrations. There have been many reported examples where protein-ligand association constants (K_a) determined using the ESI-MS assay agree with values obtained by other analytical methods.[17-23] In some instances, depending on the nature of the interactions and the experimental conditions employed, the binding measurements may be affected by the occurrence of false positives (nonspecific ligand-protein binding during the ESI process) or false negatives (gas

phase dissociation of the complex ions). However, a number of effective strategies that minimize these artifacts have been developed recently.[24-26]

The ESI-MS assay for protein-ligand affinity measurements has a number of attractive features that make it a valuable addition to the arsenal of available binding assays. The technique is fast (measurements normally can be completed in <1 min), is easily automated and consumes very little sample (typically <pmol per analysis). Additionally, there is no requirement for labeling or immobilization of the protein or the ligand, which makes the assay extremely versatile. Another attractive feature of the assay is that it allows for the simultaneous measurement of multiple binding equilibria. Consequently, the assay would seem to be well suited to library screening.

The use of ESI-MS to directly screen libraries of compounds for specific protein-ligand interactions is not new and a number of examples have been reported.[27-31] The earliest reports by Smith and coworkers,[28,29] describe the application of ESI-MS, combined with a "catch and release" (CaR) strategy, implemented with Fourier-transform ion cyclotron resonance (FTICR) MS to estimate the relative affinities of mixtures of ligands for carbonic anhydrase II (CA). Direct ESI-MS analysis of solutions containing CA and a library of benzenesulfonamides or peptides was used to detect the strongest binding ligands present in the mixture. Positive identification of the bound ligands was achieved by isolating the complex ions, followed by collision-induced (CID) dissociation and high-resolution product ion measurement. Marshall and coworkers utilized a similar CaR-ESI-MS approach to screen a library of 324 peptides against the Hck Src homology 2.[30] From their measurements they were able

to successfully identify the highest affinity polar peptide ligands, although some discrimination against hydrophobic ligands was noted.[30]

The CaR-ESI-MS approach also holds tremendous promise for screening carbohydrate libraries against carbohydrate-binding proteins to rapidly identify and quantify specific interactions.[27,31] In fact, Cederkvist *et al*. employed such an approach to screen heterochitooligosaccharides, obtained by enzymatic hydrolysis of chitosan, against chitinolytic enzyme chitinase B.[31] Based on their results, the authors suggested that the assay could be used to identify the highest affinity protein-carbohydrate interactions present in mixtures. However, given that neither the exact composition of oligosaccharide mixture used in this study, nor the affinities of the detected interactions were determined, the reliability of the assay for screening carbohydrate libraries against carbohydrate-binding proteins was not conclusively established.

Here, we describe the first detailed investigation into the application of a CaR-ESI-MS approach for screening libraries of carbohydrates against target proteins. Proof-of-concept experiments were performed to test the reliability of direct ESI-MS measurements to identify and quantify specific protein-carbohydrate interactions of varying strengths (10^3 to 10^6 M^{-1}) within mixtures of carbohydrates. The ability to positively identify isomeric ligands, following their release from protein-carbohydrate complexes in the gas phase, using CID and ion mobility separation (IMS) was also assessed.

5.2 Experimental

5.2.1 Proteins and ligands

A single chain fragment (scFv, MW 26 539 Da) of the monoclonal antibody (mAb) Se155-4 was produced using recombinant technology, as described elsewhere.[32] The antigen binding fragment (Fab, MW 48 263 Da) of the mAb CS35 was produced and purified as described previously.[33] *Clostridium difficile* toxin B subfragment (TcdB-B3, MW 30 360 Da) was expressed in *Escherichia coli* and purified as described previously.[34,35] Lysozyme (Lyz, MW 14 310 Da), ubiquitin (Ubi, MW 8 560 Da) and α-lactalbumin (αLA, MW 14 210 Da), which served as a reference protein (P_{ref}), were purchased from Sigma–Aldrich Canada (Oakville, ON). Each protein was concentrated and dialyzed against aqueous 50 mM ammonium acetate using microconcentrators (Millipore Corp., Bedford, MA) with a MW cut-off of 10 kDa and stored at –20°C if not used immediately. The carbohydrate library used in this study consisted of 209 compounds ranging in size from two to twenty-two sugar units. A complete list of the carbohydrate structures is given in Table 5.1. Stock solutions of all the carbohydrates were prepared by dissolving the solid compounds in ultrafiltered water (Milli-Q, Millipore) at a concentration of 1 mM. These were stored at -20 °C until needed.

Table 5.1 Composition of 209 component carbohydrate library.

Code	Chemical Structure	MW (Da)
L1 [a]	Methylα-D-Tal-(1-2)-[α-D-Abe-(1-3)]-α-D-Man.	486.19
L2 [a]	Methylα-D-Abe-(1-3)-2-*O*-methyl-α-D-Man-(1-3)-α-D-Glc-(1-4)-β-D-Glc.	662.26

L3 [a]	Methylα-D-Glc-(1-4)-β-D-Glc-(1-4)-[α-D-Abe-(1-3)]-α-D-Man-(1-2)-α-D- Man.	810.30
L4 [b]	Methyl β-D-Ara-(1-2)-α-D-Ara-(1-5)-[α-D-Ara-(1-3)]-α-D-Ara-(1-5)-α-D-Ara	692.62
L5 [b]	Methyl β-D-Ara-(1-2)-α-D-Ara-(1-3)-[α-D-Ara-(1-5)]-α-D-Ara-(1-5)-α-D-Ara	692.62
L6 [c]	8-Aminooctyl-methylthio-5-deoxy-a-D-Xylo-(1-4)-α-D-Man-(1-5)- α-D-Ara-(1-2)- α-D-Ara-(1-5)-[β-D-Ara-(1-2)- α-D-Ara -(1-3)]- α-D-Ara-(1-5)- α-D-Ara	1262.23
L7 [c]	8-Trifluroacetamidyl-octylβ-D-Ara-(1-2)- α-D-Ara-(1-5)- α-D-Ara-(1-5)- α-D-Ara.	1034.07
L8 [c]	8-Trifluroacetamidyl-octylβ-D-Ara-(1-2)- α-D-Ara-(1-3)-[β-D-Ara-(1-2)-α-D-Ara-(1-5)]- α-D Ara -(1-5)- α-D-Ara.	770.72
L9 [a,b]	8-Trifluroacetamidyl-octylα-D-Ara-(1-5)- α-D-Ara-(1-5)- α-D-Ara	638.64
L10 [a,b,c]	8-Trifluroacetamidyl-octyl α-D-Ara-(1-3)-[α-D-Ara-(1-5)]- α-D-Ara-(1-5)-α-D-Ara	770.77
L11 [a,b]	8-Trifluroacetamidyl-octyl α-D-Ara-(1-2)-D-Ara-(1-3)-α-D-Ara-(1-5)-α-D-Ara	770.77
L12 [a,b,c]	*p*-Methoxyphenyl 2″,3″,4″-tri-*O*-methyl-α-L-Fucp-(1-3)-α-L-Rhap-(1-3)-2-*O*-methyl-α-L-Rhap	619.32
L13 [a,b]	*p*-Methoxyphenyl 2″,3″,4″-tri-*O*-methyl-α-L-Fucp-(1-3)-2′,4′-di-*O*-benzyl-α-L-Rhap-(1-3)-4-*O*-benzyl-α-L-Rhap	605.20

L14 [a,b,c]	*p*-Methoxyphenyl 2″,4″-di-*O*-methyl-α-L-Fucp-(1-3)-α-L-Rhap-(1-3)-2-*O*-methyl-α-L-Rhap.	605.20
L15 [a,b,c]	*p*-Methoxyphenyl 3″,6″-di-*O*-methyl-β-D-Glc-(1-4)-2′,3′-di-*O*-methyl-α-L-Rhap-(1-2)-3-*O*-methyl-α-L-Rhap	649.35
L16 [a,b,c]	*p*-Methoxyphenyl 3″,6″-di-*O*-methyl-β-D-Glc-(1-4)-3′-*O*-methyl -α-L-Rhap-(1-2)-3-*O*-methyl-α-L-Rhap	621.22
L17 [a,b]	8-Aminooctyl 5-*O*-[α-D-Ara-(1-3)-[α-D-Ara-(1-5)]-α-D-Ara D-Ara-(1-5)-α-D-Ara-(1-5)-α-D-Ara-(1-3)-[5-*O*-[α-D-Ara-(3)-[α-D Ara-(1-5)]-α-D-Ara]-α-D-Ara-(1-5)-α-D-Ara-(1-5)- D-Ara-(1-5)]-α-D-Ara-(1-5)-α-D-Ara-(1-5)-α-D-Ara-(1-5)- Ara-(1-5)-α-D-Ara-(1-5)-α-D-Ara	2551.55
L18 [a,b]	8-Azidooctyl 5-*O*-{α-D-Ara-(1-5)-α-D-Ara-(1-5)-α-D-Ara-(1-5)-α-D-Ara-(1-5)-α-D-Ara-(1-3)-[α-D-Ara-(1-5)-α-D-Ara-(1-5)-α-D-Ara-(1-5)-α-D-Ara-(1-5)-α-D-Ara-(1-5)]}-α-D-Ara-(1-5)-α-D Ara-(1-5)-α-D-Ara-(1-5)-α-D-Ara-(1-5)-α-D-Ara-(1-5)]-α-D-Ara	2287.29
L19 [a ,b]	8-Azidooctyl β-D-Ara-(1-2)-α-D-Ara-(1-3)-α-D-Ara-(1-5)-α-D-Ara-(1-5)-α-D-Ara-α-D-Ara-(1-5)-α-D-Ara-(1-5)-α-D-Ara	1229.23
L20 [a,b]	Octyl α-D-Ara-(1-5)-[β-D-Gal-(1-5)-β-D-Gal-(1-6)]-β-D-Gal	749.75

L21 [a,b]	Octyl β-D-Gal-(1-5)-β-D-Gal-(1-4)-α-L-Rhap-(1-3)-2-acetamido-2-deoxy-α-D-Glc	804.80
L22 [a,b]	Octyl β-D-Gal-(1-5)-β-D-Gal-(1-6)-β-D-Gal	617.32
L23 [a,b]	Octyl β-D-Gal-(1-6)-β-D-Gal-(1-5)-β-D-Gal	617.32
L24 [a,b]	Methyl β-D-Ara-(1-2)-α-D-Ara	296.30
L25 [a,b]	Methyl β-D-Ara-(1-2)-α-D-Ara-(1-5)-α-D-Ara	428.43
L26 [a,b]	Methyl β-D-Ara-(1-2)-α-D-Ara-(1-3)-α-D-Ara	428.43
L27 [a,b]	Methyl β-D-Ara-(1-2)-α-D-Ara-(1-3)-α-D-Ara-(1-5)-α-D-Ara	560.56
L28 [a,b]	Methyl β-D-Ara-(1-2)-α-D-Ara-(1-5)-[α-D-Ara-(1-3)-]α-D-Ara	560.56
L29 [a,b]	Methyl β-D-Ara-(1-2)-α-D-Ara-(1-3)-[α-D-Ara-(1-5)-]α-D-Ara	560.56
L30 [a,b]	Methyl α-D-Ara-(1-5)-α-D-Ara-(1-5)-α-D-Ara	428.43
L31 [a,b]	Methyl α-D-Ara-(1-5)-[α-D-Ara-(1-3)-]α-D-Ara-(1-5)-α-D-Ara	560.56
L32 [a,b]	Methyl α-D-Ara-(1-5)-[α-D-Ara-(1-3)-]α-D-Arab	428.43
L33 [a,b]	Methyl α-D-Ara-(1-3)-α-D-Ara	312.31
L34 [a,b]	Methyl α-D-Ara-(1-3)-α-D-Ara-(1-5)-α-D-Ara	444.44
L35 [a,b]	β-D-Gal(1-4)β-D-Glc	342.34
L36 [a,b,c]	α-D-Fuc(1-2)β-D-Gal(1-4)β-D-Glc	488.49

L37 [a,b]	α-D-Fuc(1-2)β-D-Galβ(1-4)β-D-Gal	488.49
L38 [a,b]	β-D-Galβ(1-4)[α-D-Fuc(1-3)]β-D-Glc	488.49
L39 [a,b]	β-D-Galβ(1-4)[α-D-Fuc(1-3)]β-D-Glc	488.49
L40 [a,b,c]	β-D-Gal(1-3)β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc	707.71
L41 [a,b]	β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc	545.55
L42 [a,b]	β-D-Gal(1-4)β-D-GlcNAc(1-3)β-D-Galβ(1-4)β-D-Glc	707.71
L43 [a,b]	β-D-Gal(1-4)β-D-GlcNAc(1-6)β-D-Gal[β-D-(1-4)Glc](1-3)β-D-GlcNAc(1-4)β-D-Gal	1074.07
L44 [a,b]	β-D-Gal(1-4)β-D-GlcNAcβ(1-3)β-D-Gal(1-4)β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc	1073.07
L45 [a,b]	α-D-Fucα(1-2)β-D-Gal(1-4)[α-D-Fuc(1-3)]β-D-Glc	634.63
L46 [a,b]	α-D-Fuc(1-2)β-D-Gal(1-3)[α-D-Fuc(1-4)]β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc	1001.00
L47 [a,b,c]	α-D-Fuc(1-2)β-D-Gal(1-3)β-D-GlcNAc(1-3)β-D-Galβ(1-4)β-D-Glc	853.85
L48 [a,b]	β-D-Gal(1-3)[α-D-Fuc(1-4)] β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc	853.85
L49 [a,b]	β-D-Gal(1-4)[α-D-Fuc(1-3)]β-D-GlcNAc(1-3) β-D-Gal(1-4)β-D-Glc	853.85
L50 [a,b]	β-D-Gal(1-4)β-D-GlcNAc(1-3)β-D-Gal(1-4)[α-D-Fuc(1-3)]β-D-Glc	853.85

L51 [a,b]	β-D-Gal(1-3)β-D-GlcNAc[α-D-Fuc(1-4)](1-3)β-D-Galβ(1-4) β-D-Glc[α-D-Fuc (1-3)]	1001.00
L52 [a,b]	β-D-Gal(1-4)[α-D-Fuc(1-3)]β-D-GlcNAc(1-6)[α-D-Fuc(1-2)β-D-Gal(β1-3)β-D-GlcNAc(β1-3)]β-D-Gal(β1-4)β-D-Glc	1366.37
L53 [a,b]	β-D-Galβ(1-3)[α-D-Fuc(1-4)]β-D-GlcNAc(1-3)β-D-Gal(1-4)[α-D-Fuc(1-3)]β-D-GlcNAc(1-3)]β-D-Gal(1-4)β-D-Glc	1366.37
L54 [a,b]	5-Acetyl-α-Neu(2-3)β-D-Gal(1-4)β-D-Glc	633.63
L55 [a,b]	5-Acetyl-α-Neu(2-6)β-D-Gal(1-4)β-D-Glc	633.63
L56 [a,b]	5-Acetyl-α-Neu(2-3)β-D-Gal(1-4)[α-D-Fuc(1-3)]β-D-Glc	779.78
L57 [a,b]	5-Acetyl-α-Neu(2-3)β-D-Gal(1-3)β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc	999.00
L58 [a,b]	5-Acetyl-α-Neu(2-6)[β-D-Gal(1-3)]β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc	999.00
L59 [a,b]	5-Acetyl-α-Neu(2-6)β-D-Gal(1-4)β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc	999.00
L60 [a,b]	5-Acetyl-α-Neu(2-3)β-D-Gal(1-4)β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc	999.00
L61 [a,b]	5-Acetyl-α-Neu(2-3)β-D-Gal(1-3)[5-Acetyl-α-Neu(2-6)]β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc	1291.29

L62 [a,b]	5-Acetyl-α-Neu(2-3)β-D-Gal(1-3)[α-D-Fuc(1-4)]β-D-GlcNAc(1-3)β-D-Galβ(1-4)β-D-Glc	1146.15
L63 [a,b]	α-D-Fuc(1-2)β-D-Galβ(1-3)[5-Acetyl-α-Neu (2-6)]β-D-GlcNAcβ(1-3)β-D-Galβ(1-4)β-D-Glc	1146.15
L64 [a,b]	α-D-GalNAc(1-3)[α-D-Fuc(1-2)]β-D-Galβ(1-3)β-D-GlcNAc	843.84
L65 [a,b]	α-D-GalNAc(1-3)[α-D-Fuc(1-2)]β-D-Gal(1-4) β-D-GlcNAc	843.84
L66 [a,b]	α-D-GalNAc(1-3)[α-D-Fuc(1-2)]β-D-Gal(1-3)β-D-GalNAc	843.84
L67 [a,b]	α-D-GalNAc(1-3)[α-D-Fuc(1-2)]β-D-Gal (1-3)β-D-GalNAc	843.84
L68 [a,b]	α-D-GalNAc(1-3)[α-D-Fuc(1-2)]β-D-Gal(1-3)β-D-Gal	801.80
L69 [a,b]	α-D-GalNAc(1-3)[α-D-Fuc (1-2)]β-D-Gal(1-4)β-D-Gal	801.80
L70 [a,b]	α-D-Gal(1-3)[α-D-Fuc(1-2)]β-D-Gal (1-3)β-D-GlcNAc	801.80
L71 [a,b]	α-D-Gal(1-3)[α-D-Fuc(1-2)]β-D-Gal(1-4)β-D-GlcNAc	801.80
L72 [a,b]	α-D-Gal(1-3)[α-D-Fuc(1-2)]β-D-Gal(1-3)β-D-GlcNAc	801.80
L73 [a,b]	α-D-Gal(1-3)[α-D-Fuc(1-2)]α-D-Gal(1-3)β-D-GlcNAc	801.80
L74 [a,b]	α-D-Gal(1-3)[α-D-Fuc(1-2)]α-D-Gal(1-3)α-D-Gal	760.76
L75 [a,b]	α-D-Gal(1-3)[α-D-Fuc(1-2)]α-D-Gal(1-4)α-D-Gal	760.76
L76 [a,b]	α-D-Fuc(1-2)α-D-Gal(1-3)β-D-GlcNAc	640.64

L77 [a,b]	α-D-Fuc(1-2)α-D-Gal(1-4)β-D-GlcNAc	640.64
L78 [a,b]	α-D-Fuc(1-2)β-D-Gal(1-3)β-D-GalNAc	640.64
L79 [a,b]	α-D-Fuc(1-2)β-D-Gal(1-3)β-D-GalNAc	640.64
L80 [a,b]	α-D-Fuc(1-2)β-D-Gal(1-3)β-D-Gal	688.69
L81 [a,b]	α-D-Fuc(1-2)β-D-Gal(1-4)β-D-Gal	598.60
L82 [a,b]	α-D-GalNAc(1-3)[α-D-Fuc(1-2)]β-D-Gal	529.53
L83 [a,b]	α-D-GalNAc(1-3)[α-D-Fuc(1-2)]β-D-Gal(1-4)β-D-Glc	691.69
L84 [a,b]	α-D-GalNAc(1-3)[α-D-Fuc(1-2)]β-D-Gal (1-3)β-D-GlcNAc(1-3)β-D-Gal	894.89
L85 [a,b]	α-D-GalNAc(1-3)[α-D-Fuc(1-2)]β-D-Gal (1-4)β-D-GlcNAc(1-3)β-D-Gal	894.89
L86 [a,b]	α-D-GalNAc(1-3)[α-D-Fuc(1-2)]β-D-Gal (1-3)β-D-GalNAc(1-3)β-D-Gal	894.89
L87 [a,b]	α-D-GalNAc(1-3)[α-D-Fuc(1-2)] β-D-Gal (1-3)β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc	1057.06
L88 [a,b]	α-D-GalNAc(1-3)[α-D-Fuc(1-2)]β-D-Gal(1-4)β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Glc	1057.06
L89 [a,b]	β-D-Gal(1-3)β-D-GalNAc(1-3)β-D-Gal	545.55
L90 [a,b]	α-D-Gal(1,3)β-D-Gal(1,4)β-D-GlcNAc	546.05
L91 [a,b]	α-D-Gal(1-3)[α-D-Fuc(1-2)]β-D-Gal(1-4)β-D-Glc	650.65
L92 [a,b]	α-D-Gal(1-3)[α-D-Fuc(1-2)]β-D-Gal(1-4)β-D-GlcNAc(1-	853.85

	3)β-D-Gal	
L93 [a,b]	α-D-Gal(1-3)[α-D-Fuc(1-2)]β-D-Gal(1-3)β-D-GalNac(1-3)β-D-Gal	853.85
L94 [a,b]	α-D-Gal(1-3)[α-D-Fuc(1-2])β-D-Gal(1-3)β-D-GlcNac(1-3)-β-D-Gal(1-4)β-D-Glc	1016.02
L95 [a,b]	α-D-Gal(1-3)[α-D-Fuc(1-2)]β-D-Gal(1-4)β-D-GlcNac(1-3)-β-D-Gal(1-4)β-D-Glc	1016.02
L96 [a,b]	α-D-Fuc(1,2)β-D-Gal	326.63
L97 [a,b]	α-D-Fuc(1,3)β-D-Gal	326.33
L98 [a,b]	α-D-Fuc(1,2)β-D-Gal(1-4)β-D-GlcNAc	529.53
L99 [a,b]	α-D-Fuc(1-2)β-D-Gal(1-4)β-D-GlcNAc(1-3)β-D-Gal	691.69
L100 [a,b]	α-D-Fuc(1-2)β-D-Gal(1-4)β-D-GlcNAc(1-4)β-D-Gal	691.69
L101 [a,b]	α-D-Fuc(1-2)β-D-Gal(1-3)β-D-GlcNAc(1-3)β-D-Gal	691.69
L102 [a,b]	α-D-Fuc(1-2)β-D-Gal(1-3)β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Gal	853.85
L103 [a,b]	α-D-Fuc(1-2)β-D-Gal(1-4)β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-Gal	853.85
L104 [a,b]	β-D-Gal(1-3)β-D-GlcNAc	383.38
L105 [a,b]	β-D-Gal(1-3)[α-D-Fuc(1-4)]β-D-GlcNAc	529.53
L106 [a,b]	β-D-Gal(1-3)[α-D-Fuc(1-4)]β-D-GlcNAc(1-3)β-D-Gal	691.69
L107 [a,b]	β-D-Gal(1-3)[α-D-Fuc(1-4)]β-D-GlcNAc(1-3)β-D-	1000.00

	Galβ(1-4)β-D-Gal[α-D-Fuc(1-3)]β-D-Glc	
L108 [a,b]	α-D-Fuc(1,2)β-D-Gal(1,3)[α-D-Fuc(1,4)]β-D-GlcNAc	675.68
L109 [a,b]	α-D-Fuc(1,2)β-D-Gal(1,3)[α-D-Fuc(1,4)]β-D-GlcNAc(1-3)β-D-Gal	837.84
L110 [a,b]	β-D-Gal(1,4)[α-D-Fuc(1,3)]β-D-GlcNAc	529.53
L111 [a,b]	β-D-Gal(1,4)[α-D-Fuc(1,3)]β-D-GlcNAc(1-3)β-D-Gal	691.69
L112 [a,b]	5-Acetyl-α-Neu(2-3)β-D-Gal(1-4)[α-D-Fuc(1-3)]β-D-GlcNAc GlcNAc(1-3) β-D-Gal	982.98
L113 [a,b]	β-D-Gal(1,4)[α-D-Fuc(1,3)]β-D-GlcNAc(1-3)β-D-Gal(1-4)[α-D-Fuc(1-3)]β-D-Glc	1000.00
L114 [a,b]	α-D-Fuc(1,2)β-D-Gal(1,4)[α-D-Fuc(1,3)]β-D-GlcNAc	676.68
L115 [a,b]	α-D-Fuc(1,2)β-D-Gal(1,4)[α-D-Fuc(1,3)]β-D-GlcNAc(1-3)β-D-Gal	837.84
L116 [a,b]	α-D-Gal(1-4)β-D-Gal(1-4)β-D-Glc	504.50
L117 [a,b]	β-D-GlcNAc(1-3)α-D-Gal(1-4)β-D-Gal(1-4)β-D-Glc	707.71
L118 [a,b]	5-Acetyl-α-Neu(2-3)β-D-Gal(1-3)β-D-GalNAc(1-4)[5-Acetyl-α-Neu (2-3)]β-D-Gal(1-4)β-D-Glc	1290.29
L119 [a,b]	β-D-Gal(1-3)β-D-GalNAc(1-4)[5-Acetyl-α-Neu(2-8)5-Acetyl-α-Neu(2-3)]β-D-Gal(1-4) β-D-Glc	1290.29
L120 [a,b]	β-D-GalNAc(1-4)[5-Acetyl-α-Neu(2-8)5-Acetyl-α-Neu(2-3)]β-D-Gal(1-4)β-D-Glc	1128.13

L121 [a,b]	5-Acetyl-α-Neu(2-8)5-Acetyl-α-Neu(2-3)β-D-Gal(1-4)β-D-Glc	924.92
L122 [a,b]	β-D-Gal(1-3)β-D-GalNAc(1-4)[5-Acetyl-α-Neu(2-3)]β-D-Gal(1-4)β-D-Glc	999.00
L123 [a,b]	5-Acetyl-α-Neu(2-3)β-D-Gal(1-3)β-D-GalNAc(1-4)β-D-Gal(1-4)β-D-Glc	999.00
L124 [a,b]	α-D-Fuc(1-2)β-D-Gal(1-3)β-D-GalNAc(1-4)[5-Acetyl-α-Neu(2-3)]β-D-Gal(1-4)β-D-Glc	1145.14
L125 [a,b]	β-D-Gal(1-3)β-D-GalNAc(1-4)β-D-Gal(1-4)β-D-Glc	707.71
L126 [a,b]	β-D-GalNAc(1-4)[5-Acetyl-α-Neu(2-3)]β-D-Gal(1-4)β-D-Glc	836.84
L127 [a,b]	β-D-GalNA(1-4)β-D-Gal(1-4)β-D-Glc	545.55
L128 [a,b]	5-Acetyl-α-Neu(2-8)5-Acetyl-α-Neu(2-3)β-D-Gal(1-3)β-D-GalNAc(1-4)[5-Acetyl-α-Neu(2-3)]β-D-Gal(1-4)β-D-Glc	1581.58
L129 [a,b]	β-D-Gal(1-3)β-D-GalNAc(1-4)[5-Acetyl-α-Neu(2-8)5-Acetyl-α-Neu (2-8)5-Acetyl-α-Neu(2-3)]β-D-Gal(1-4)β-D-Glc	1581.58
L130 [a,b]	β-D-GalNAc(1-4)[5-Acetyl-α-Neu(2-8)5-Acetyl-α-Neu(2-8)5-Acetyl-α-Neu(2-3)]β-D-Gal(1-4)β-D-Glc	1419.42
L131 [a,b]	5-Acetyl-α-Neu(2-8)5-Acetyl-α-Neu(2-8)5-Acetyl-α-Neu(2-3)β-D-Gal(1-4)β-D-Glc	1216.22

L132 [a,b]	β-D-Glc(1-4)β-D-Glc (1-4)β-D-Glc	342.34
L133 [a,b,c]	β-D-Glc(1-4)β-D-Glc (1-4)β-D-Glc	504.50
L134 [a,b,c]	β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc	667.67
L135 [a,b,c]	β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc	827.83
L136 [a,b,c]	β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc	990.99
L137 [a,b,c]	β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc	1153.15
L138 [a,b,c]	β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc	1315.31
L139 [a,b,c]	α-D-Man(1-6)[α-D-Man(1-3)]α-D-Man(1-6)[α-D-Man(1-3)]α-D-Man	829.83
L140 [a,b,c]	β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc	667.67
L141 [a,b,c]	β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc(1-4)β-D-Glc	990.99
L142 [a,b]	α-D-GalNAc(1-3)[α-D-Fuc(1-2)]β-D-Gal(1-3)β-D-GalNAc(1-3)α-D-Gal (1-4)β-D-Galβ(1-4)β-D-Glc	1219.22
L143 [a,b]	α-D-Gal(1-3)[α-D-Fuc(1-2)]β-D-Gal(1-3)β-D-GalNAc(1-3)α-D-Gal (1-4) β-D-Gal(1-4)β-D-Glc	1178.18
L144 [a,b]	α-D-Fuc(1-2)]β-D-Gal(1-3)β-D-GalNAc(1-3)α-D-Gal (1-	1016.02

	4) β-D-Gal(1-4)β-D-Glc	
L145 [a,b]	β-D-Gal(1-3)β-D-GalNAc(1-3)α-D-Gal (1-4) β-D-Gal(1-4)β-D-Glc	869.87
L146 [a,b]	β-D-GalNAc(1-3)α-D-Gal(1-4)β-D-Gal(1-4)β-D-Glc	707.71
L147 [a,b]	β-D-Gal(1-3)β-D-GalNAc(1-3)α-D-Gal(1-3)β-D-Galβ(1-4)β-D-Glc	869.87
L148 [a,b]	α-D-Gal (1-3)α-D-Gal(1-4)β-D-GalNAc(1-3)β-D-Galβ(1-4)β-D-Glc	869.87
L149 [a,b]	α-D-Gal (1-3)α-D-Gal(1-4)β-D-Glc	504.50
L150 [a,b]	α-D-GalNAc(1-3)β-D-GalNAc(1-3)α-D-Gal(1-3)α-D-Gal(1-4)β-D-Glc	910.91
L151 [a,b]	5-Acetyl-α-Neu(2-3)β-D-Gal(1-3)GalNAcβ(1-3) α-D-Gal(1-4)β-D-Gal(1-4)β-D-Glc	1161.16
L152 [a,b]	α-D-GalNAc(1-3)β-D-GalNAc(1-3)α-D-Gal	586.59
L153 [a,b]	5-Acetyl-α-Neu(2-3)β-D-Gal(1-3)β-D-GalNAc(1-3)α-D-Gal	836.84
L154 [a,b]	α-D-GalNAc(1-3)β-D-GalNAc(1-3)α-D-Gal(1-4)α-D-Gal(1-4)β-D-Glc	910.91
L155 [a,b]	α-D-Gal(1-4)α-D-Gal(1-4)β-D-GlcNAc	545.55
L156 [a,b]	α-D-GalNAc(1-3)β-D-Glc	383.38
L157 [a,b]	α-D-Gal(1-3)-[α-L-Fuc (1-6)]β-D-GlcNAc	529.53

L158 [a,b]	α-D-Tal(1-3)-β-D-Glc	342.34
L159 [a,b]	α-D-TalNAc(1-3)-β-D-Glc	383.38
L160 [a,b]	α-D-GalNAc(1-3)β-D-Glc	383.38
L161 [a,b]	α-D-Gal(1-4)β-D-GlcNAc	415.42
L162 [a,b]	Methylβ-D-Gal(1-4)-β-D-Glc	356.36
L163 [a,b,c]	4,6-O-benzylidene-α-D-Glc(1,4)-α-D-Glc	444.44
L164 [a,b]	β-D-Gal(1-3)[α-D-Fuc(1-6)]β-D-GlcNAc	529.53
L165 [a,b]	Trichloroethanolβ-D-GlcNAc(1-4)-β-D-GlcNAc	556.56
L166 [a,b]	α-D-Fuc(1-2)β-D-Gal(1-3)β-D-GlcNAc	415.42
L167 [a,b]	β-D-Gal(1-4)α-D-Glc	400.40
L168 [a,b]	α-D-Tal(1-3)β-D-Glc	342.34
L169 [a]	α-D-GlcNAc(1-3)β-D-Glc	383.38
L170 [a,b]	Octanol-α-D-Ara(1-5)(2-3)-anhydro-α-D-Ara	390.39
L171 [a,b]	Aminohexanol-β-D-Gal(1-4)β-D-GlcNAc(1-3)β-D-Gal(1-4)β-D-GlcNAc acetic acid salt	848.85
L172 [a,b]	Aminohexanol β-D-Gal(1-3) β-D-GlcNAc(1-3) β-D-Gal(1-4) β-D-GlcNAc acetic acid salt	848.85
L173 [a,b]	Aminohexanol β-D-Gal(1-4)[α-D-Fuc(1-3)] β-D-GlcNAc(1-3)β-D-Gal(1,4) β-D-GlcNAcacetic acid salt	993.99
L174 [a,b]	Aminohexanol α-D-Gal(1-3) β-D-Gal(1-4)[α-D-Fuc(1-3)] β-D-GlcNAc acetic acid salt	851.85

L175 [a,b]	Aminohexanol β-D-Galβ(1-4)[α-D-Fucα(1-3)])β-D-GlcNAcβ acetic acid salt	689.69
L176 [a,b]	Aminohexanol β-D-Gal(1-4) β-D-GlcNAc(1-3)β-D-Gal (1-4)[α-D-Fuc (1-3)]β-D-GlcNAc acetic acid salt	993.99
L177 [a,b]	6-Azidoβ-D-Gal(1-4)β-D-GlcNAc(1-3)β-D-Gal(1-4)[α-D-Fuc (1-3)]β-D-GlcNAc ethanol	964.96
L178 [a,b]	6-Azido α-D-Gal (1-3)β-D-Gal (1-4)[α-D-Fuc(1-3)]β-D-GlcNAc	761.76
L179 [a,b]	6-Azidoβ-D-Gal(1-3)[α-D-Fuc(1-4)]β-D-GalNAc(1-3)β-D-Gal(1-4)[α-D-Fuc(1-3)]β-D-GlcNAc	1111.11
L180 [a,b]	6-Azido β-D-Gal(1-3)β-D-GlcNAc(1-3)β-D-Galβ(1-4)β-D-GlcNAc	818.82
L181 [a,b]	6-Azidoβ-D-Gal(1-4)[α-D-Fuc(1-3)]β-D-GlcNAc	599.60
L182 [a,b]	β-D-Gal(1-3)β-D-GlcNAc	383.38
L183 [a,b]	β-D-GlcNAc(1-4) β-D-GlcNAc(1-4) β-D-GlcNAc(1-4)β-D-GlcNAc	831.83
L184 [a,b]	β-D-Gal (1-6) β-D-Gal	342.34
L185 [a,b]	α-D-Gal(1-3)β-D-Gal (1-4)α-D-Gal(1-3)β-D-Gal	667.67
L186 [a,b]	α-D-Gal(1-3)β-D-Gal(1-4)α-D-Gal	649.65
L187 [a,b]	α-D-Gal(1-3)β-D-Gal	342.34
L188 [a,b]	α-L-Fuc(1-6)β-D-GlcNAc	365.37

L189 [a,b]	α-D-GlcNAc(1-4)β-D-Gal	383.38
L190 [a,b,c]	α-D-Man(1-6)[α-D-Man (1-3)]α-D-Man (1-6)α-D-Man	667.67
L191 [a,b]	N-Acetyl-D-lactosamine	383.38
L192 [a,b]	N,N',N'',N''',N'''',N'''''-Hexaacetyl chitohexaose	1238.24
L193 [a,b]	Methyl α-D-Abe	148.34
L194 [a,b]	β-D-Gal-(1-4)-[α-L-Fuc-(1-3)]-β-D-GlcNAc-(1-3)-β-D-Gal-(1-4)-[α-L-Fuc-(1-3)]-β-D-GlcNAc-(1-*O*)-(6-aminohexanol acetic acid salt)	1150.98
L195 [a,b]	β-D-GlcNAc-(1-3)-β-D-Gal-(1-4)-β-D-Glc-(1-*O*)-(6-azidohexanol)	671.62
L196 [a,b]	β-D-GlcNAc-(1-3)-β-D-Gal-(1-4)-[α-L-Fuc-(1-3)]-β-D-GlcNAc-(1-*O*)-(6-aminohexanol acetic acid salt)	850.59
L197 [a,b]	β-D-GlcNAc-(1-3)-β-D-Gal-(1-4)-[α-L-Fuc-(1-3)]-β-D-GlcNAc-(1-*O*)-(2-aminoethanol acetic acid salt)	836.46
L198 [a,b]	β-D-GlcNAc-(1-3)-β-D-Gal-(1-4)-β-D-GlcNAc-(1-*O*)-(6-aminohexanolaceticacid salt)	686.47
L199 [a,b]	β-D-Gal-(1-4)-β-D-GlcNAc-(1-3)-β-D-Gal-(1-4)-β-D-Glc-(1-*O*)-(6 aminohexanol acetic acid salt)	822.12
L200 [a,b]	α-D-Man	180.25
L201 [a,b]	α-D-Gal	180.25
L202 [a,b]	Methyl α-D-Glc	194.28
L203 [a,b]	Methyl α-D-Gal	356.49

L204 [a,b]	β-D-Gal(1-4)β-D-GlcNAc(1-3) β-D-Gal(1-4)GlcNAc(1-3)β-D-Gal(1-4)β-D-GlcNAc(1-3) β-D-Gal(1-4) β-D-Glc	1556.72
L205 [a,b]	α-D-Gal(1-3)(α-D-Fuc(1-2)β-D-Gal(1-3)β-D-GlcNAc(1-3)β-D-Gal	853.28
L206 [a,b]	β-D-Gal(1-3)β-D-GlcNAc(1-3)β-D-Gal(1-4)[α-D-Fuc(1-3]β-D-Glc	853.28
L207 [a,b]	5-Acetyl-α-Neu(2-3) β-D-Gal(1-3)β-D-GlcNAc(1-3)β-D-Gal	836.42
L208 [a,b]	5-Acetyl-α-Neu(2-3)β-D-Gal(1-4)β-D-GlcNAc(1-3)β-D-Gal	836.42
L209 [a,b]	β-D-GalNAc(1-4)[5-Acetyl-α-Neu(2-3)]β-D-Gal(1-4)β-D-Glc	835.42

a. Carbohydrates contained in library screened against scFv.

b. Carbohydrates contained in library screened against FAB CS35.

c. Carbohydrates contained in library screened againstTcdB-B3.

Compounds **L1-3, L156-170, L193** were donated by Prof. David Bundle (University of Alberta).

Compounds **L4-34**, **L64-81** were donated by Prof. Todd Lowary (University of Alberta).[36-41]

Compounds **L171-181, L194-199** were donated by Prof. Kenneth Ng (University of Calgary).

Compounds **L37**, **L39**, **L41**, **L42**, **L44**, **L50, L54, L57, L60, L82-89, L91-95, L97, L99-102, L106, L107, L109, L125-131, L142-154, L204-209** were purchased from Elicityl SA (Crolles, France).

Compounds **L35, L200-202** were purchased from Sigma–Aldrich Canada (Oakville, ON).

Compounds **L43, L51, L90, L96, L98, L103-105, L108, L110, L114, L132-141, L155, L182-192** were purchased from Dextra (Reading, UK).

Compounds **L36, L38, L40, L45-49, L52, L53, L55, L56, L58, L59, L61-63** were purchased from IsoSep AB (Sweden).

5.2.2 Mass spectrometry

All experiments were carried out using 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. Mass spectra were obtained in either positive or negative ion modes using cesium iodide (concentration 30 ng μL^{-1}) for calibration. Given below are instrumental conditions used for measurements carried out in positive ion mode. For negative ion mode measurements, the polarities were switched. To perform nanoESI, tips were produced from borosilicate capillaries (1.0 mm o.d., 0.68 mm i.d.) pulled to ~5 μm using a P-97 micropipette puller (Sutter Instruments, Novato, CA). A platinum wire was inserted into the nanoESI tip and a capillary voltage of 1.0-1.3 kV was applied to carry out ESI. A cone voltage of 50-90 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, ion-mobility, and transfer ion guides, were maintained at 20 V, 28 V, and

10 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. The helium chamber preceding the traveling wave ion mobility (TWIMS) device was maintained at 7.72 mbar. All traveling-wave ion mobility measurements were carried out using N_2 as the mobility gas at a pressure of 1.88 mbar. A linear ion mobility T-wave was used with a fixed wave height of 8.5 V and velocity of 250 m s^{-1}. Data acquisition and processing were carried out using MassLynx (v 4.1).

To confirm the identity of carbohydrate "hits" in the ESI mass spectra, ions corresponding to a specific protein-carbohydrate complex were isolated using the quadrupole mass filter and then subjected to CID, by increasing the injection voltage (30-65 V) into the trap ion guide, to release the ligand. Ligands ejected from the complex in an ionized form could then be subjected to MS analysis, allowing for a more accurate MW determination. The released ligands could also be subjected to IMS and the corresponding arrival time distribution (ATD) compared to reference ATDs measured for all carbohydrates in the library. Following IMS, ligands could also be fragmented by increasing the injection voltage into the transfer ion guide and the resulting CID mass spectra compared to reference CID mass spectra measured for all carbohydrates in the library.

5.3 Data analysis

5.3.1 Establishing absolute K_a values.

As described in detail elsewhere,[17,24] in cases where a single protein (P) and ligand (L) are present in solution, the ratio (*R*) of the total ion abundance (*Ab*) of ligand-bound protein to free protein ions (eq 1.3) measured by ESI-MS for solutions of known initial

concentrations of protein ($[P]_o$) and ligand ($[L]_o$) can be used to calculate association constant K_a, eq 1.2.

In cases where the solution contains multiple ligands (L_1, L_2, …., L_X) absolute K_a values for individual ligands, i.e., $K_{a,Lx}$, can be found using eq 5.1:

$$K_{a,Lx} = \frac{[PL_x]}{[P][L_x]} = \frac{R_{PLx}}{[L_x]} \tag{5.1}$$

where $[PL_x]$ is the concentration of protein bound to the ligand L_x, [P] and $[L_x]$ are the concentrations of free P and L_x, respectively, at equilibrium and R_{PLx} is the ratio of the total ion abundance of protein bound to L_x and free protein. The value of $[L_x]$ can be found using the equation of mass balance:

$$[L_x] = [L_x]_o - [PL_x] \tag{5.2}$$

where $[PL_x]$ is calculated from $[P]_o$ and the individual R_{PLx} values, eq 5.3:

$$[PL_x] = \frac{[P]_o R_{PLx}}{1 + R_{PL1} + ... + R_{PLx}} \tag{5.3}$$

The value of $K_{a,Lx}$ can then be calculated using eq 5.4:

$$K_{a,Lx} = \frac{R_{PLx}}{[L_x]_o - \frac{[P]_o R_{PLx}}{(1 + R_{PL1} + ... R_{PLx})}} \tag{5.4}$$

In the case where P possesses multiple (N) ligand binding sites, the apparent (macroscopic) association constant ($K_{a,Lx,i}$) corresponding to the attachment of i (= 1, 2, …, N) molecules of L_x can be determined using eq 5.5:

$$K_{a,Lx,i} = \frac{[P(L_x)_i]}{[P(L_x)_{i-1}][L_x]} = \frac{R_{PLx,i}}{R_{PLx,(i-1)}[L_x]} \tag{5.5}$$

where $R_{\mathrm{PLx},i}$ and $R_{\mathrm{PLx},(i-1)}$ are the ratios of the total ion abundance of ligand-bound protein (to i or i-1 molecules of L_x) and free protein. The free ligand concentration at equilibrium, $[L_x]$, can be found from the equation of mass balance:

$$[L_x] = [L_x]_o - \sum_i i[P(L_x)_i] \tag{5.6}$$

and the values of $[P(L_x)_i]$ can be calculated from eq 5.7:

$$[P(L_x)_i] = \frac{[P]_o R_{P(Lx)i}}{1 + \sum_{x,i} R_{P(Lx)i}} \tag{5.7}$$

5.3.2 Correction for nonspecific protein-carbohydrate binding.

In order to correct ESI mass spectra for the occurrence of nonspecific carbohydrate-protein binding during the ESI process, the reference protein method was employed.[24] This method involves the use of a reference protein (P_{ref}), which does not bind specifically to the target protein or the carbohydrates. Because the distributions of nonspecifically bound ligand can vary between measurements,[24,42,43] P_{ref} must be present in the solution being analyzed. The distributions of carbohydrates bound nonspecifically to P_{ref} are used to quantitatively correct the mass spectrum for the occurrence of nonspecific carbohydrate binding to the target protein and any specific protein-carbohydrate complexes present in solution. A detailed description of the P_{ref} method has been given elsewhere.[24] An overview of application of the method in the case of multiple (distinct) ligands binding to a target protein with multiple binding sites is given below.

The "true" abundance of P and a given $P(L_x)_i$ species in solution (i.e., $Ab(P)$ and $Ab(P(L_x)_i)$ can be related to the apparent (measured) abundance, $Ab_{app}(P)$ and $Ab_{app}(P(L_x)_i)$, as shown in eqs 5.8a and 5.8b:

$$Ab_{app}(\mathrm{P}) = f_0 Ab(\mathrm{P}) \tag{5.8a}$$

$$Ab_{app}(\mathrm{P(L_x)}_i\ ) = f_0 Ab(\mathrm{P(L_x)}_i) + \sum_{0<k\le i} f_k Ab\mathrm{P(L_x)}_{i-k} \tag{5.8b}$$

where f_0 is the fractional abundance of $\mathrm{P(L_x)}_i$ that does not undergo nonspecific binding to $\mathrm{L_x}$ during the ESI process and f_k is the fractional abundance of $\mathrm{P(L_x)}_{i\text{-}k}$ that binds nonspecifically to k molecules of $\mathrm{L_x}$ to produce $\mathrm{P(L_x)}_i$. Taking the ratio of eqs 5.8a and 5.8b gives eq 5.9:

$$\frac{Ab_{app}(\mathrm{P(L_x)}_i)}{Ab_{app}(\mathrm{P})} = \frac{Ab(\mathrm{P(L_x)}_i)}{Ab(\mathrm{P})} + \sum_{0<k\le i} \frac{f_k Ab(\mathrm{P(L_x)}_{i-k})}{f_0 Ab(\mathrm{P})} \equiv R_{\mathrm{P(Lx)}i} + \sum_{0<k\le i} \frac{f_k}{f_0} R_{\mathrm{P(Lx)}(i\,-\,k)} \tag{5.9}$$

The ratio of f_k/f_0 is found using eq 5.10:

$$f_k/f_0 = Ab(\mathrm{P_{ref}(L_x)}_k)/Ab(\mathrm{P_{ref}}) \tag{5.10}$$

where $Ab(\mathrm{P_{ref}})$ and $Ab(\mathrm{P_{ref}(L_x)}_k)$ are the measured abundances of ions corresponding to free $\mathrm{P_{ref}}$ and $\mathrm{P_{ref}}$ bound nonspecifically to k molecules of $\mathrm{L_x}$. Rearrangement of eq 5.9 gives the following expression for $R_{\mathrm{P(Lx)}i}$:

$$R_{\mathrm{P(Lx)}i} = \frac{Ab_{app}(\mathrm{P(L_x)}_i)}{Ab_{app}(\mathrm{P})} - \sum_{0<k\le i} \frac{f_k}{f_0} R_{\mathrm{P(Lx)}(i\,-\,k)} \tag{5.11}$$

It is important to note that the $R_{\mathrm{P(Lx)}(i\text{-}k)}$ terms in eq 5.11 are based on the "true" abundances , i.e., abundances that have been corrected for nonspecific binding. Once the magnitude of $R_{\mathrm{P(Lx)}i}$ has been established, the corresponding $\mathrm{K_{a,Lx,}}_i$ can be calculated using eq 5.4 or 5.5.

A critical assumption underlying the $\mathrm{P_{ref}}$ method is that, in a given ESI-MS measurement, the occurrence of nonspecific carbohydrate (ligand) binding is independent of the nature of the proteins present. The reliability of the method has

been rigorously tested for ESI-MS binding measurements performed using FTICR MS instruments.[24,42,43] However, to our knowledge, the method has not been tested on the Synapt G2 Q-IMS-TOF instrument employed in the present study. To ensure that the method is reliable, a series of control experiments were performed. ESI mass spectra were measured for solutions containing a pair of proteins (Lyz with scFv or Ubi) and a non-interacting carbohydrate, maltotriose (**L133**) or maltopentaose (**L135**), and the distributions of carbohydrate bound nonspecifically to the proteins were compared. Shown in Figures 5.1 and 5.2 are representative ESI mass spectra acquired in positive ion mode for solutions of **L133** and **L135** and each protein pair. Inspection of the ESI mass spectra reveals signals corresponding to protonated protein ions, as well as ions corresponding to protein bound to as many as three carbohydrate molecules, i.e., $(\text{scFv} + q\text{L})^{n+}$ where $q = 0 - 3$. Also shown in Figures 5.1 and 5.2 are the normalized distributions of carbohydrate bound nonspecifically to each protein. Importantly, for a given protein pair, the distributions are identical within experimental error. These results confirm that P_{ref} method can be used to correct ESI mass spectra, acquired with the Synapt G2 Q-IMS-TOF mass spectrometer, for the occurrence of nonspecific carbohydrate-protein binding.

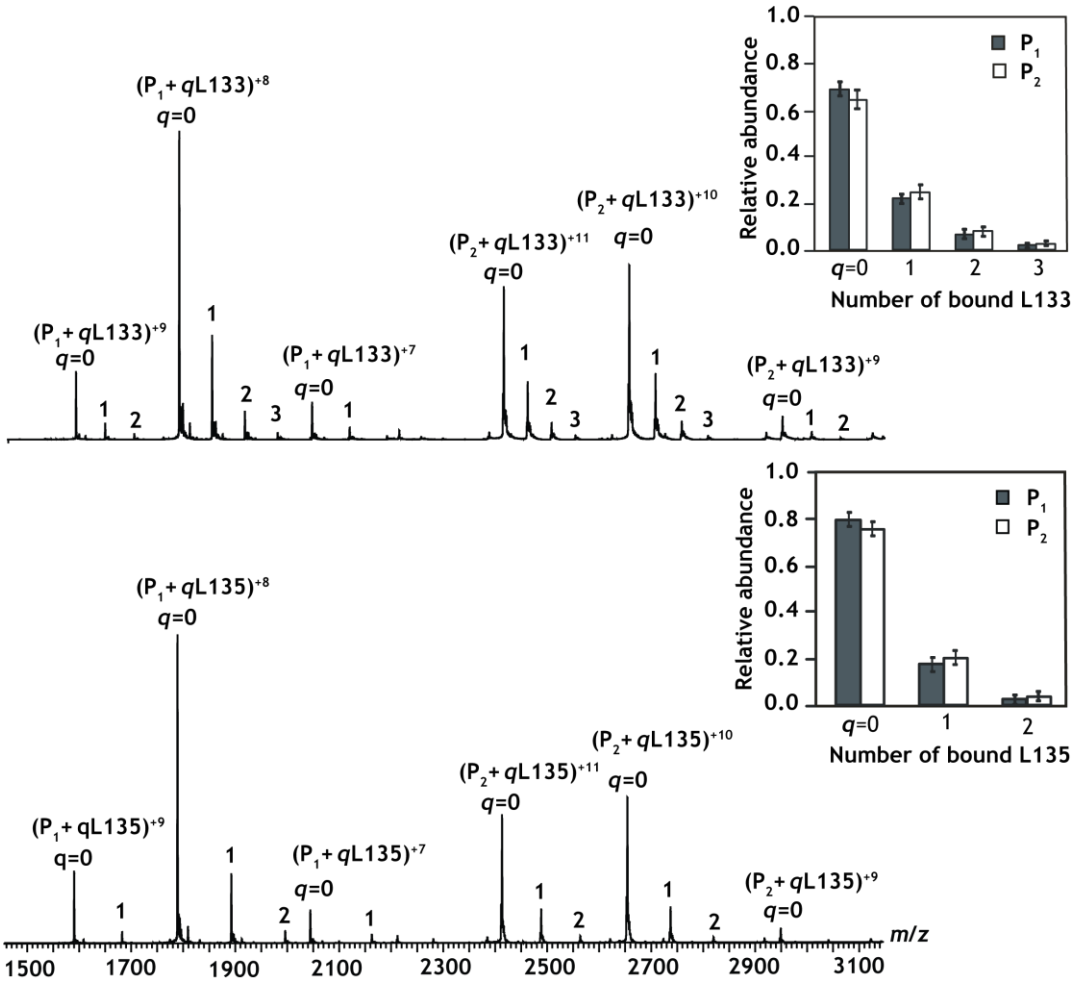


Figure 5.1 Representative ESI mass spectra acquired in positive ion mode for solutions of Lyz (P_1, 10 μM), scFv (P_2, 10 μM) and (a) **L133** (100 μM) and (b) **L135** (100 μM)**.** Also shown are the normalized distributions of **L133** and **L135** bound nonspecifically to P_1 and P_2, as determined from the mass spectra. The reported errors correspond to 2 standard deviations.

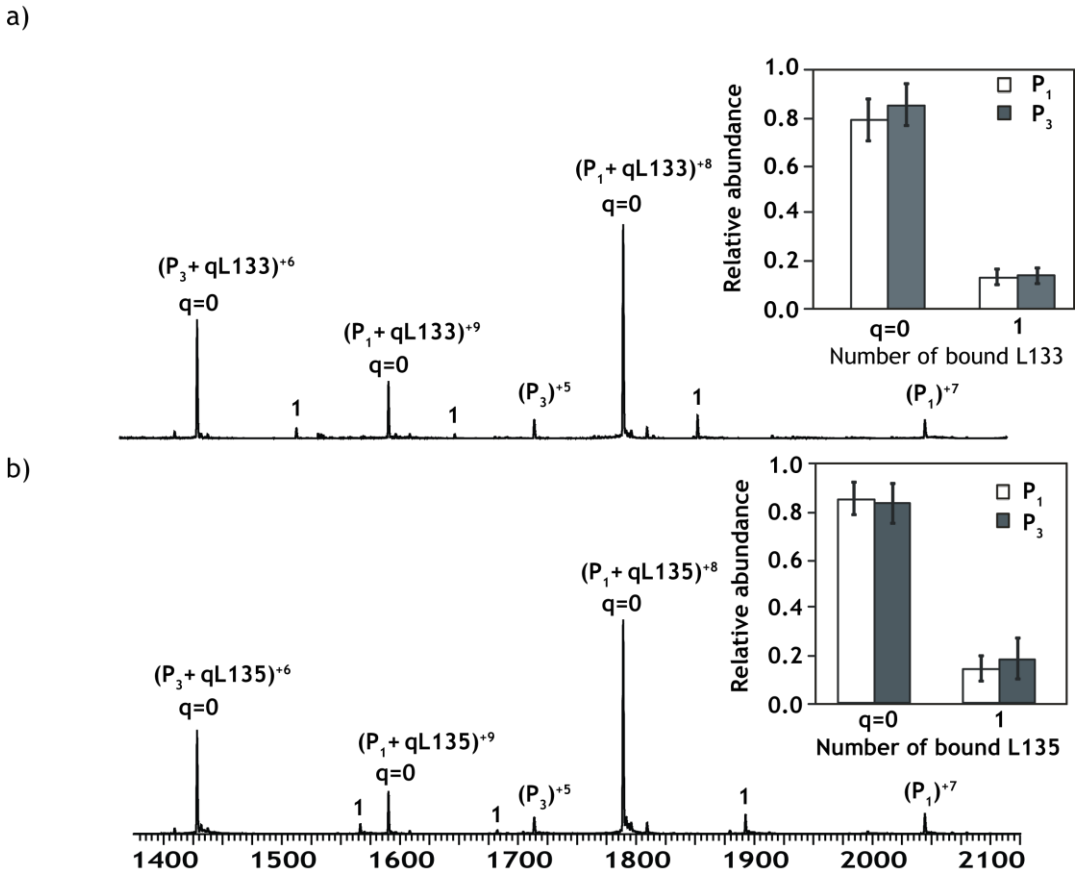


Figure 5.2 Representative ESI mass spectra acquired in positive ion mode for solutions of Lyz (P_1, 7.5 μM), Ubi (P_3, 6 μM) and (a) **L133** (100 μM) and (b) **L135** (100 μM)**.** Also shown are the normalized distributions of **L133** and **L135** bound nonspecifically to P_1 and P_3, as determined from the mass spectra. The reported errors correspond to 2 standard deviations.

5.4 Results and discussion

5.4.1 Screening carbohydrate libraries for moderate-to-high affinity interactions

In order to test the utility of direct ESI-MS measurements for identifying protein-carbohydrate interactions of moderate affinity (i.e., $10^4 - 10^6$ M^{-1}) within a mixture of carbohydrates, a number of control experiments were performed on solutions containing Se155-4 scFv and a library of 204 carbohydrates. The carbohydrate binding properties of the antibody have been extensively characterized and the binding epitope established.[32,38] The components of the carbohydrate library used in these experiments were carefully selected so that only three moderate affinity ligands were present, a tri- (**L1**), tetra- (**L2**), and pentasaccharide (**L3**). The affinities of these ligands for scFv ($(5.7\pm0.4)x10^4$ (**L1**), $(1.6\pm0.1)x10^5$ (**L2**) and $(3.4\pm0.1)x10^5$ M^{-1} (**L3**)) were determined individually using the direct ESI-MS assay in separate experiments. Shown in Figure 5.3 are representative ESI mass spectra acquired in positive (Figure 5.3a) and negative ion mode (Figure 5.3b) for a solution containing scFv (5 μM) and the carbohydrate library (1 μM each). Due to the relatively low concentration of the library components, significant nonspecific binding of the carbohydrates to the scFv was not anticipated and, consequently, no P_{ref} was added to the solution for these measurements. In addition to abundant signal corresponding to free carbohydrate ions, signals corresponding to free scFv and scFv bound to **L1**, **L2** and **L3** were evident in the mass spectra. Absolute affinities were calculated from the mass spectra following the method described in the Experimental section. The affinities obtained from the measurements in positive ion mode ($3.4x10^4$ M^{-1} (**L1**), $1.7x10^5$ M^{-1} (**L2**) and $1.2x10^5$ M^{-1} (**L3**)) and negative ion mode ($3.8x10^4$ M^{-1} (**L1**), $2.1x10^5$ M^{-1} (**L2**) and $2.2x10^5$ M^{-1}

(**L3**)) are similar in magnitude. Notably, the absolute affinities determined for **L1** and **L2** from the ESI-MS analysis of the library agree within a factor of two with the values determined from individual measurements; while the values for **L3** are within a factor of 3. Therefore, from a single ESI-MS measurement performed on a library of >200 carbohydrates, the three specific protein-carbohydrate interactions were successfully identified and quantified.

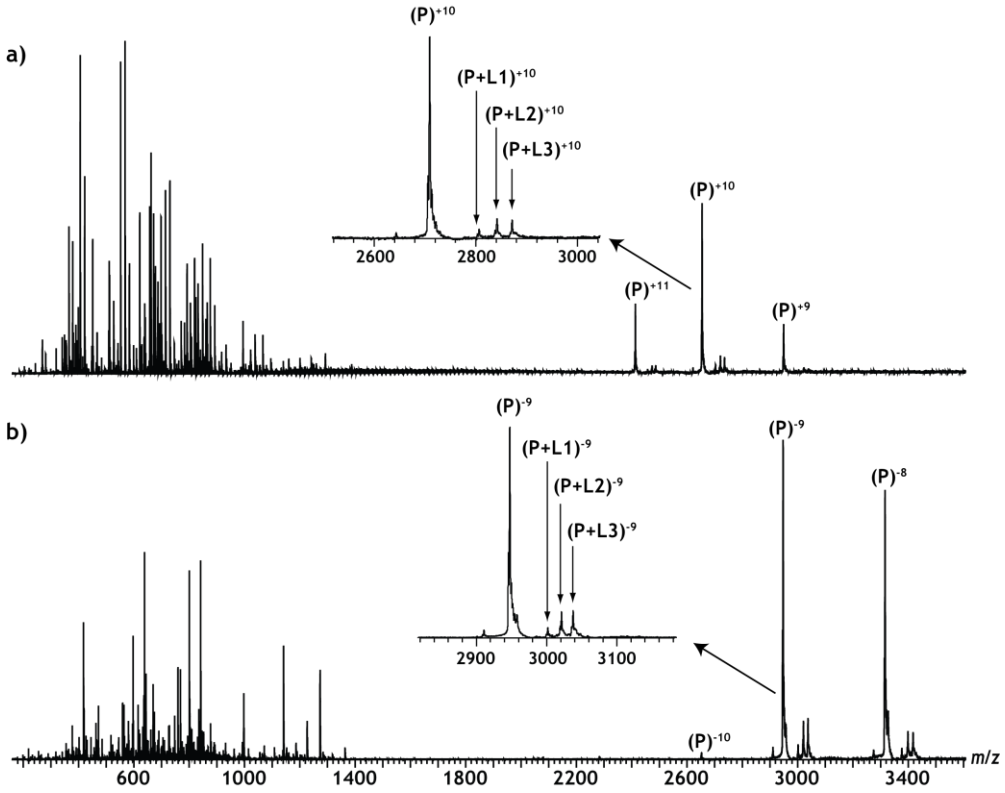


Figure 5.3 Representative ESI mass spectra obtained for a solution of scFv (P, 5 μM) and 204 carbohydrates (1 μM each) acquired in (a) positive ion mode and (b) negative ion mode.

Two important considerations when designing a library screening experiment are the concentrations of the library components and the concentration ratio of the protein to the library components. The use of low concentrations for the library components is desirable for a number of reasons. First, it reduces the probability of nonspecific carbohydrate binding to the target protein during the ESI process. Secondly, it reduces the suppression of the protein ion signal (by the components of the library) and, at the same time, allows for the analysis of larger libraries. Thirdly, it reduces the consumption of compounds that may be expensive to purchase or difficult to synthesize. However, low ligand concentrations hamper the detection of low affinity protein-ligand interactions. This effect can be mitigated, to some extent, by reducing the protein concentration, which increases the fraction of ligand-bound protein in solution. In order to establish optimal concentrations for the target protein and the library components for the successful detection of specific protein-carbohydrate interactions, a series of ESI-MS measurements were performed on solutions containing different concentrations of scFv and library. Shown in Figures 5.4 and 5.5 are ESI mass spectra acquired in positive and negative ion modes, respectively, for solutions containing scFv (5 μM) and the library components at concentrations of 1.00, 0.25, 0.10 and 0.01 μM. It can be seen that, in both positive and negative ion mode, ion signal corresponding to the three specific (scFv + L) complexes was detected at library concentrations as low as 0.25 μM. Under these conditions, each of the specific complexes account for between 0.7% and 2.2% of the scFv in solution. At 0.1 μM only the two highest affinity interactions were detected, while at 0.01 μM only free protein was detected. Improved detection of the specific ligands at low library concentrations

was achieved by decreasing the concentration of protein. Shown in Figure 5.6 is an illustrative ESI mass spectra acquired in positive ion mode for solutions with a fixed scFv and library molar concentration ratio of 5:1 and absolute scFv concentrations of 5, 2.5, 1.25 and 0.625 μM. It can be seen that the three specific (scFv + L) complexes are more clearly evident at the reduced scFv concentrations. Furthermore, the absolute affinities are in good agreement with the values obtained at higher concentrations. For example, analysis of the mass spectrum acquired in positive ion mode for the solution containing scFv (1.25 μM) and library (0.25 μM) yields affinities of 5.1×10^4 M^{-1} (**L1**), 2.0×10^5 M^{-1} (**L2**) and 1.6×10^5 M^{-1} (**L3**).

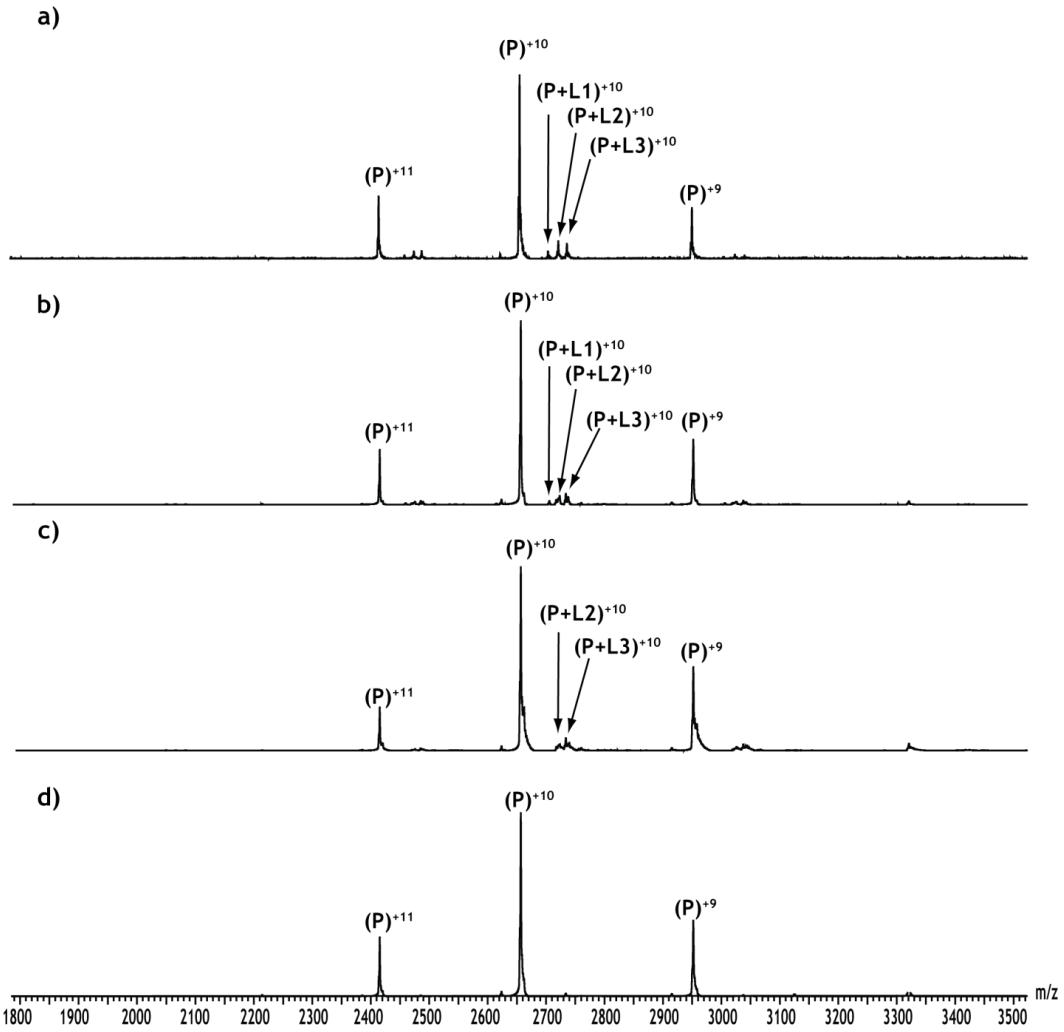


Figure 5.4 Representative ESI mass spectra acquired in positive ion mode for aqueous ammonium acetate solutions (50 mM, pH 7) containing 5 μM scFv (P) and (a) 1 μM, (b) 0.25 μM, (c) 0.1 μM and (d) 0.01 μM of the 204 carbohydrate library. The library contains three moderate affinity ligands, **L1**, **L2**, and **L3**.

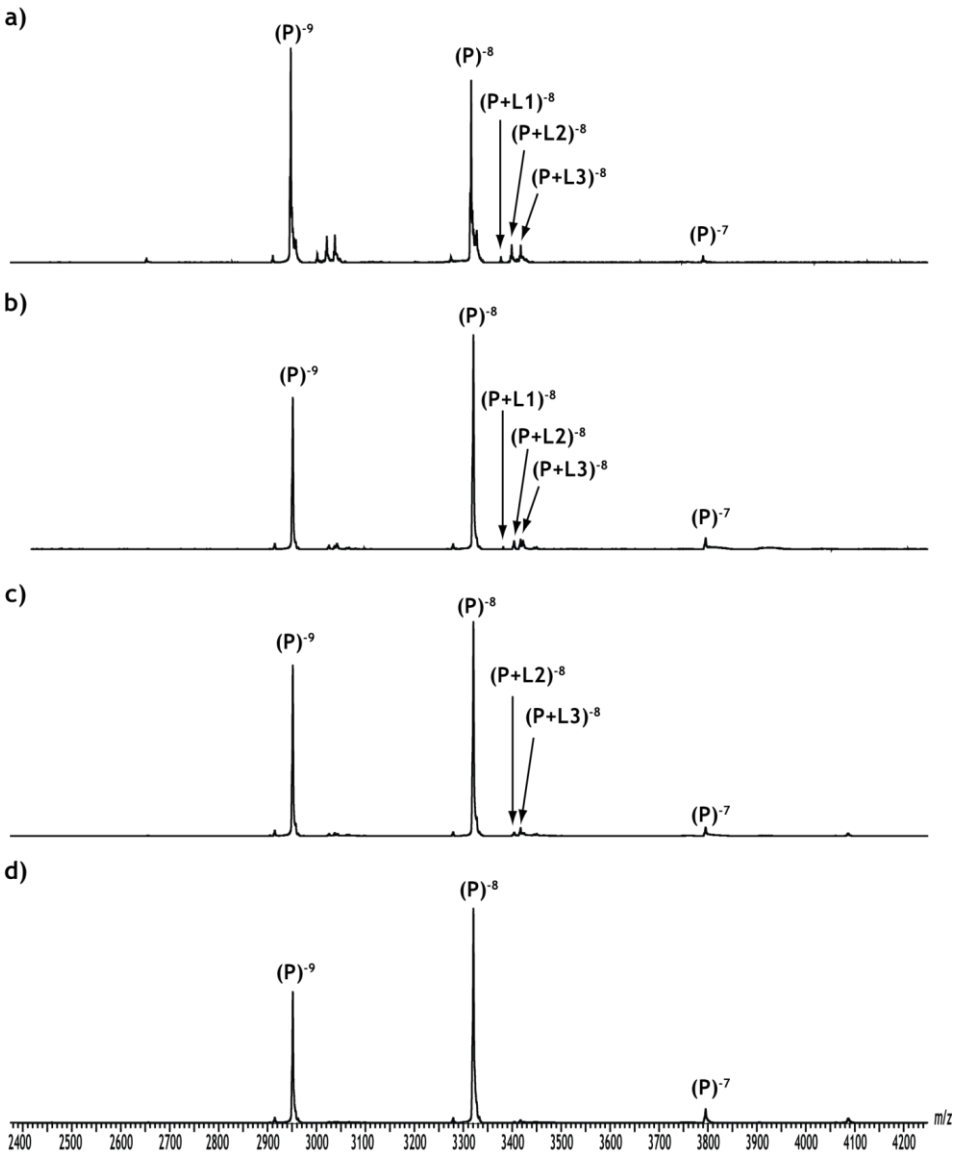


Figure 5.5 Representative ESI mass spectra acquired in negative ion mode for aqueous ammonium acetate solutions (50 mM, pH 7) containing 5 μM scFv (P) and (a) 1 μM, (b) 0.25 μM, (c) 0.1 μM and (d) 0.01 μM of the 204 carbohydrate library. The library contains three moderate affinity ligands, **L1**, **L2**, and **L3**.

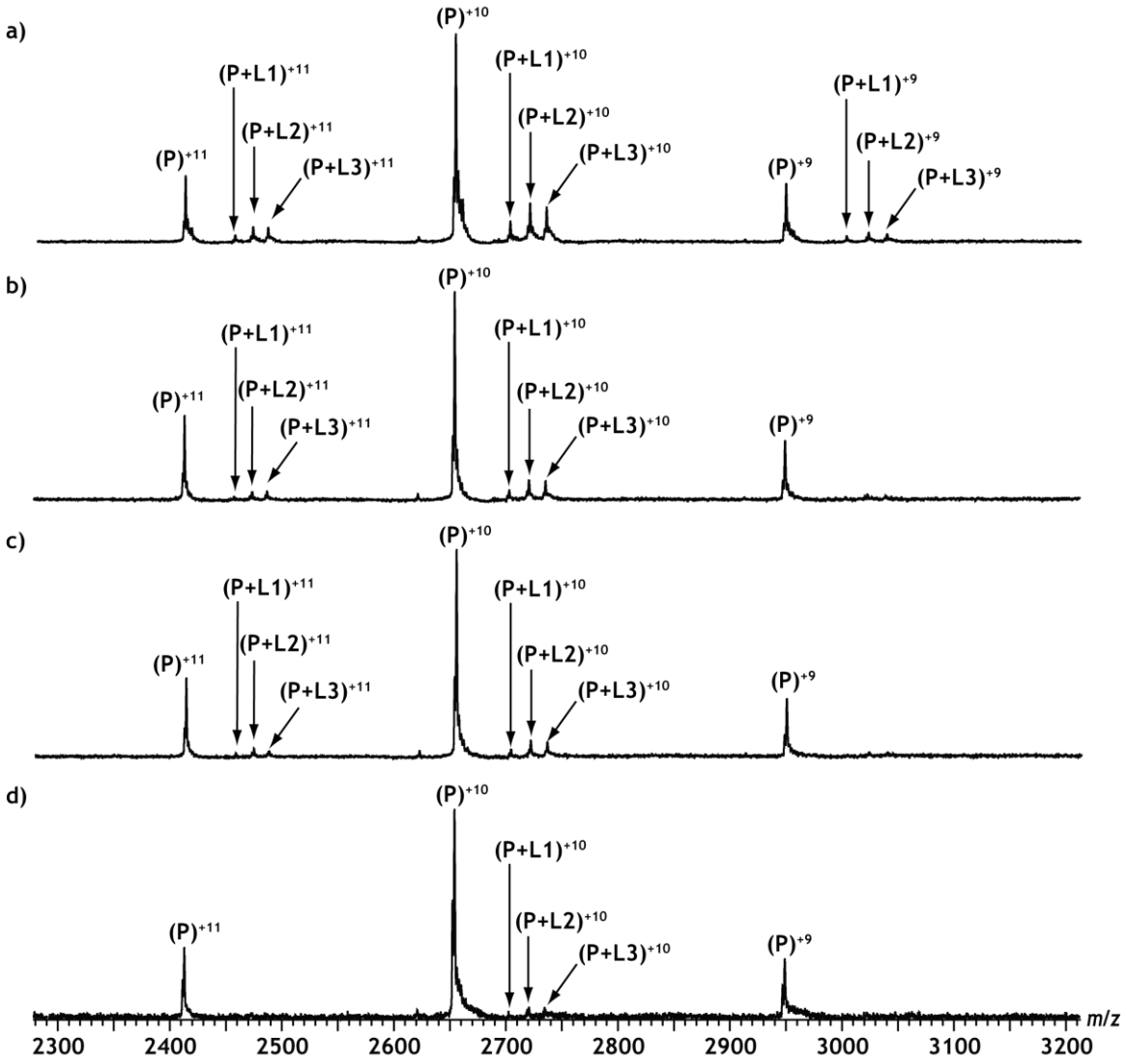


Figure 5.6 Representative ESI mass spectra obtained in positive ion mode for aqueous ammonium acetate (50 mM) solutions containing scFv (P) and 204 carbohydrates at different concentrations: (a) 5 μM scFv and 1 μM of each carbohydrate, (b) 2.5 μM scFv and 0.5 μM of each carbohydrate, (c) 1.25 μM scFv and 0.25 μM of each carbohydrate and (d) 0.625 μM scFv and 0.125 μM of each carbohydrate. The library contains three moderate affinity ligands, **L1**, **L2**, and **L3**.

Given that the MWs of **L1**, **L2** and **L3** are unique within the library of carbohydrates, it was possible to positively identify them based on the MWs of the corresponding (scFv + L) complexes, as determined from the ESI mass spectrum. Nevertheless, these interactions served as useful model systems for establishing conditions for the release of carbohydrate ligands, in an ionized form, and to assess the effectiveness of CID and IMS analysis of the released ligands for identification purposes. In separate experiments, the $(scFv + L)^{+10}$ and $(scFv + L)^{-9}$ ions were isolated using the quadrupole mass filter and subjected to CID, over a range of collision energies, in the trap ion guide. For the protonated $(scFv + L)^{+10}$ ions, the loss of neutral L was readily observed but no protonated L ions were detected at any of the collision energies investigated (Figure 5.7). This observation is not unexpected given the relatively low gas phase basicities of neutral carbohydrates.[45] The appearance of ions corresponding to sodiated and potassiated L was observed in some instances. These ions originate from $(scFv + L)^{+10}$ ions containing Na^+ or K^+, which were co-isolated with the protonated $(scFv + L)^{+10}$ ions.

In contrast, collisional activation of the $(scFv + L)^{-9}$ ions led to abundant loss of deprotonated ligand for all three ligands. Shown in Figure 5.8 are representative CID mass spectra acquired for the $(scFv + L)^{-9}$ ions at a collision energy of 40 V. It can be seen that, in each case, the dominant dissociation channel involves the loss of the deprotonated ligand. That the carbohydrate ligands are preferentially released in their deprotonated form can be explained by the low intrinsic gas phase acidities (GA) of neutral carbohydrates. For example, the GA of glucose has been determined to be between 328 and 336 kcal/mol.[46] In contrast, the GAs of the carboxylic acid side

chains of Asp and Glu are ~240 kcal/mol.[47] As a result, the carbohydrate ligands can effectively compete with the protein for negative charge. Following their release, the deprotonated L were subjected to IMS and CID.

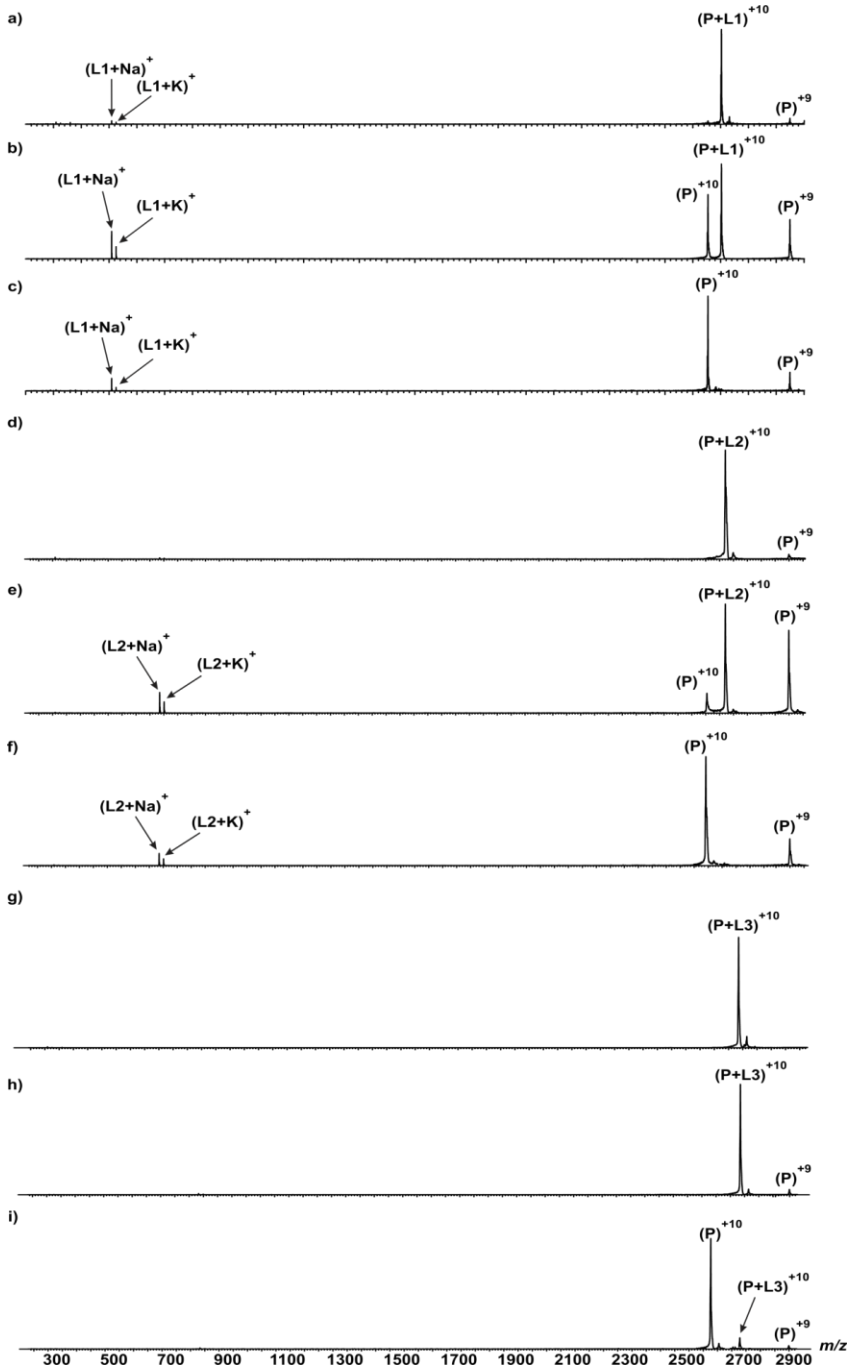


Figure 5.7 Illustrative CID mass spectra measured for (scFv + **L1**)$^{+10}$ at a trap voltage of (a) 20V, (b) 30V, (c) 40V, for (scFv + **L2**)$^{+10}$ at a trap voltage of (d) 20V, (e) 40V, (f) 60V, and for (scFv + **L3**)$^{+10}$ at a trap voltage of (g) 20V, (h) 40V, (i) 65V.

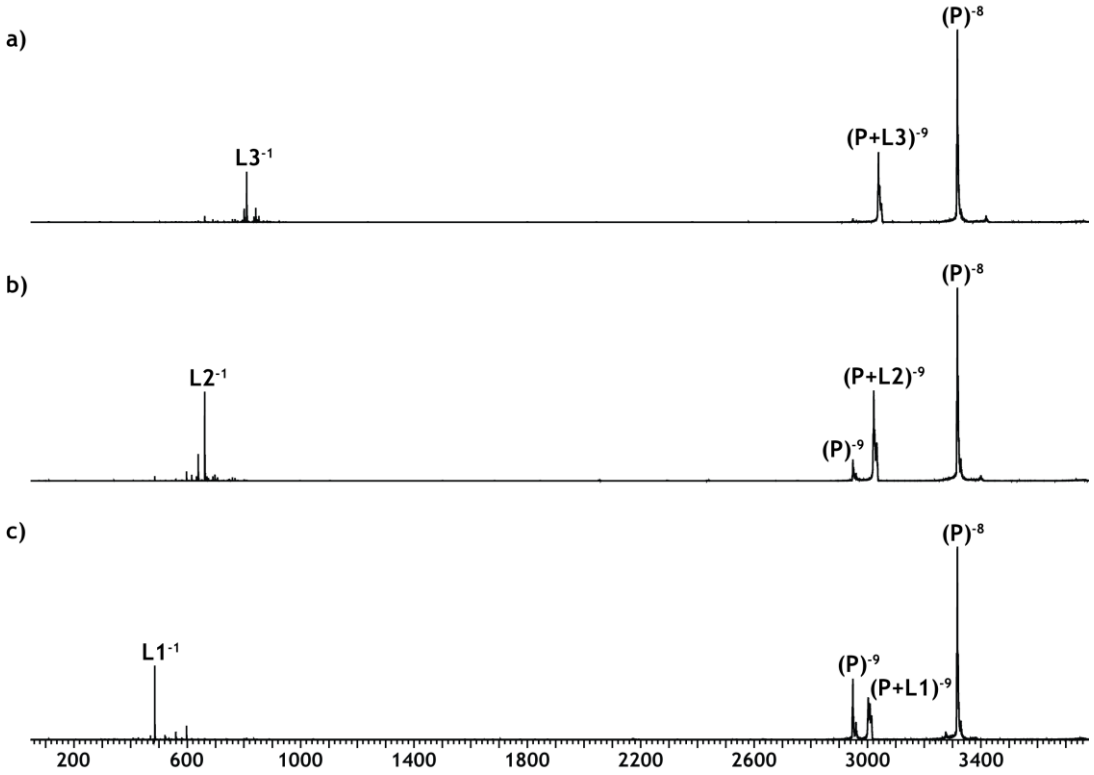


Figure 5.8 Illustrative CID mass spectra measured at a trap voltage 40 V for (a) (scFv + **L3**)$^{-9}$, (b) (scFv + **L2**)$^{-9}$, and (c) (scFv + **L1**)$^{-9}$.

Shown in Figure 5.9 are representative CID mass spectra and ATDs for deprotonated **L1**, **L2** and **L3** following release from the $(scFv + L)^{-9}$ ions (referred to as *post-release*) and the corresponding CID and ATDs measured for the deprotonated carbohydrates obtained directly from solution (referred to as *reference*). Under the IMS conditions used in the present study, the *post-release* and *reference* ATDs measured for each of the three deprotonated carbohydrates are identical, within experimental error (Figure 5.9, Table 5.2). The *post-release* and *reference* CID spectra measured for **L1** and **L2** are essentially indistinguishable. The major fragment ion observed upon CID of deprotonated **L1** is a C_2 ion (m/z 323), which results from cleavage of the glycosidic bond linking Tal and Man (Figure 5.10a).[48,49] Dissociation of deprotonated **L2** leads to the formation of a Z_3 ion (m/z 513), which corresponds to fragmentation of the glycosidic bond linking Abe and Man, and a Y_2 ion (m/z 355), which is formed from the fragmentation of the glycosidic bond between Man and Glc (Figure 5.10b). In the case of **L3**, CID results in the appearance of multiple primary and secondary fragment ions, including a B_4 (m/z 661), Z_4 (m/z 629), B_3 (m/z 485), C_2 (m/z 323), $^{0,2}X_2$ (m/z 383) and $^{0,2}X_1$ (m/z 221) ions (Figure 5.11). As was the case for **L1** and **L2**, the *post-release* and *reference* CID mass spectra for **L3** are similar in appearance although subtle differences in relative abundance of the products ions are evident.

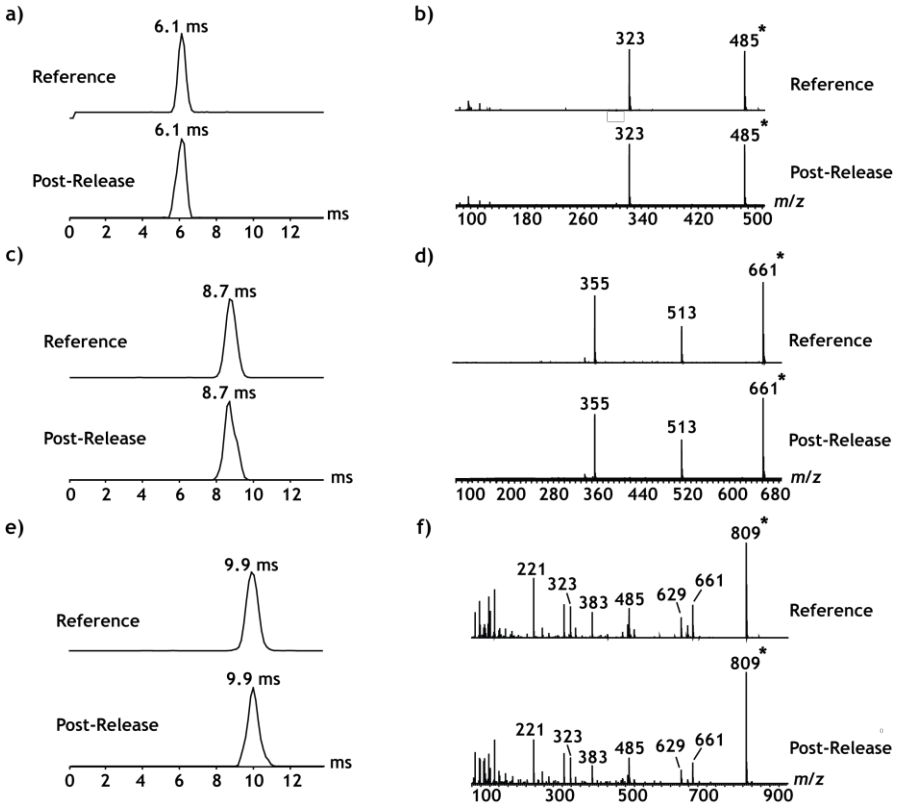


Figure 5.9 Arrival time distributions (ATDs) measured for the deprotonated ligands (L) (a) **L1** (c) **L2** and (e) **L3** following their release from the corresponding $(scFv + L)^{-9}$ ions (*post-release*). Also shown are the ATDs measured for the deprotonated **L1**, **L2** and **L3** ions obtained directly from solution (*reference*). Collision-induced dissociation (CID) mass spectra measured for the deprotonated (b) **L1** at a transfer ion guide voltage of 25V, (d) **L2** at 30V and (f) **L3** at 45V, following their release from the corresponding $(scFv + L)^{-9}$ ions (*post-release*). Also shown are the CID mass spectra measured for the deprotonated **L1**, **L2** and **L3** ions obtained directly from solution (*reference*). The deprotonated **L1**, **L2** and **L3** ions are denoted by *.

Table 5.2 Average arrival times (T_d) measured for deprotonated carbohydrate ligands (L) obtained after release from deprotonated $(P + L)^{n-}$ ions (*post-release*) or directly from solution (*reference*).[a]

Ligand	Post-release T_d (ms)	Reference T_d (ms)
L1	6.04 ± 0.07	6.09 ± 0.05
L2	8.73 ± 0.13	8.68 ± 0.05
L3	9.94 ± 0.11	9.91 ± 0.06
L4	8.03 ± 0.08	8.14 ± 0.09
L5	8.66 ± 0.05	8.55 ± 0.10

a. The reported errors correspond to one standard deviation

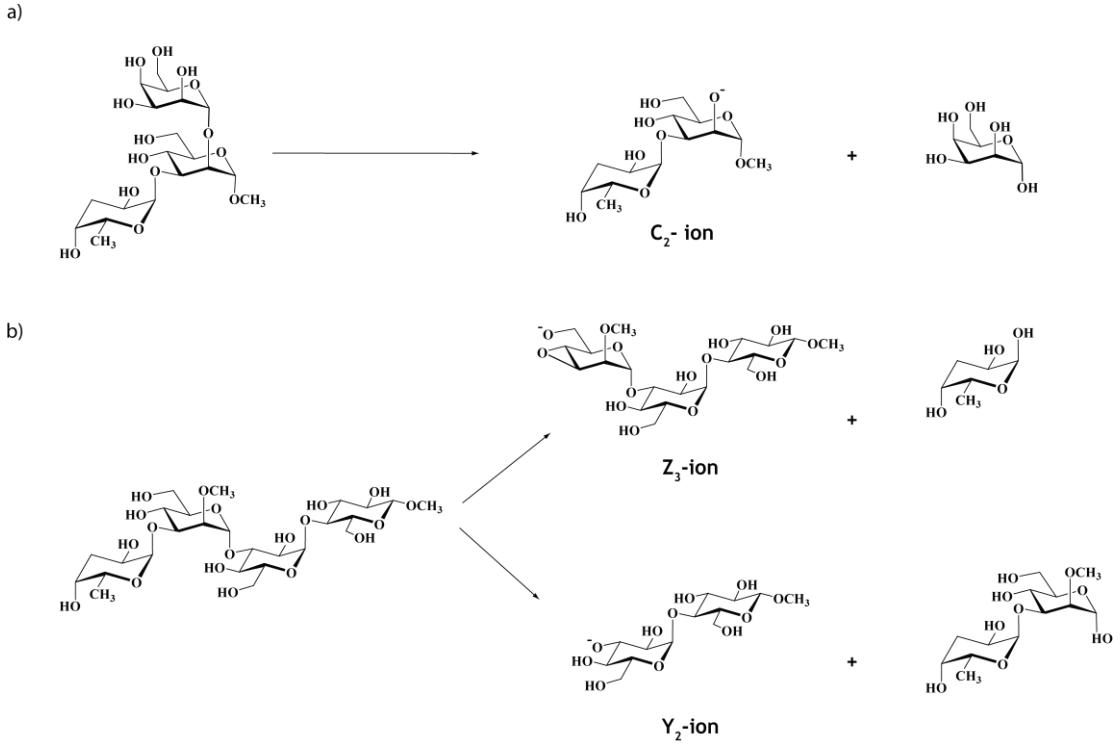


Figure 5.10 CID fragmentation pathways observed for deprotonated (a) **L1** and (b) **L2**.

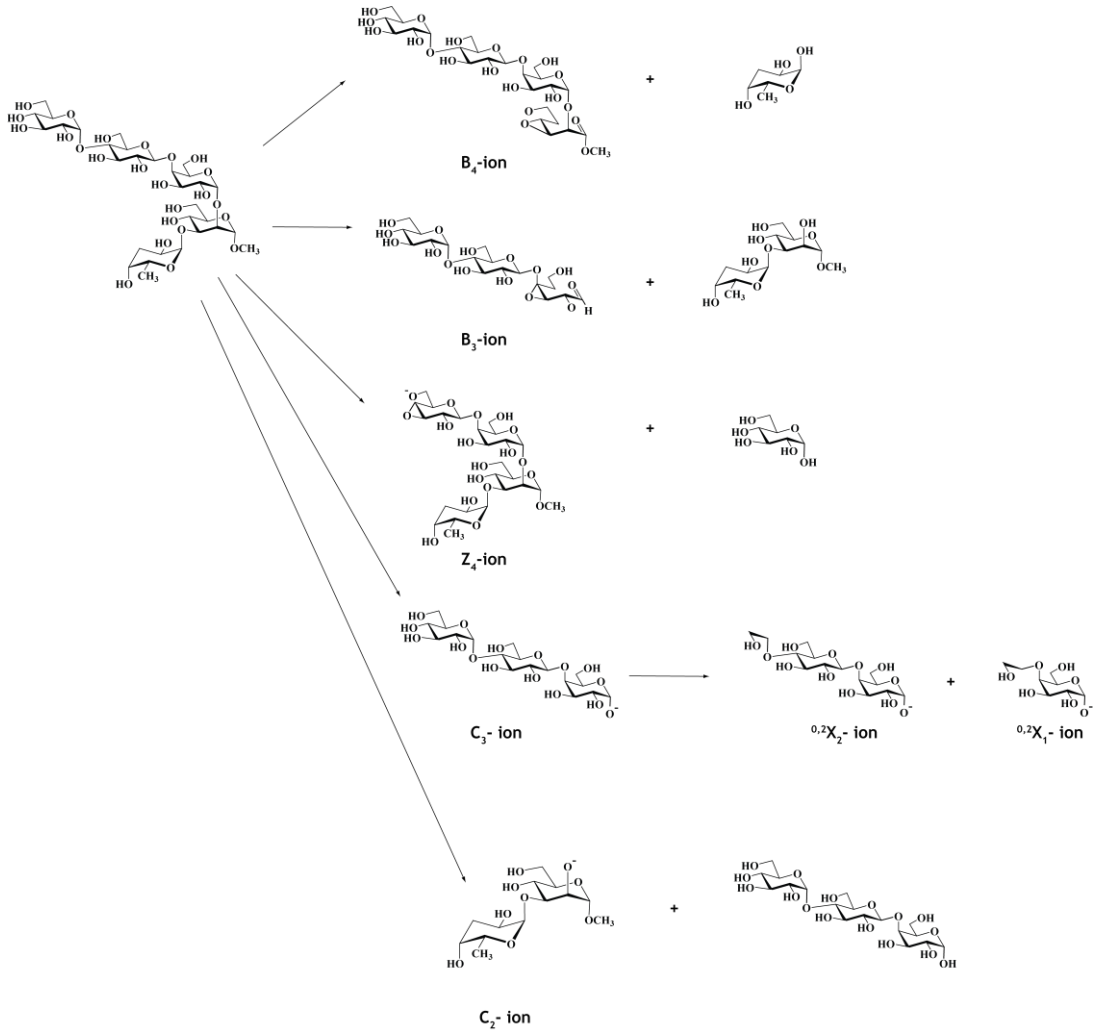


Figure 5.11 CID fragmentation pathways observed for deprotonated **L3**.

5.4.2 Distinguishing structural isomers

In the aforementioned example, the specific ligands investigated had unique MWs. However, the library of 209 carbohydrates contains numerous structural isomers. Therefore, it was of interest to test the effectiveness of the CaR-ESI-MS assay in cases where the target protein binds to isomeric ligands, such that MW alone is insufficient for ligand identification. To this end, control experiments were performed in negative ion mode on a solution of the CS35 Fab and 203 carbohydrates. Included in the library were two specific pentasaccharide ligands (**L4**, **L5**), which have identical MWs. The affinities of CS35 Fab for **L4** and **L5** were measured previously by ESI-MS to be $6.2x10^3$ and $9.9x10^4$ M^{-1}, respectively.[33] Shown in Figure 5.12a is a representative ESI mass spectrum obtained in negative ion mode for a solution of only Fab (5 μM). Analysis of the mass spectrum reveals the presence of four isoforms (referred to as P (MW 48.25 kDa), P′(MW 48.92 kDa), P′′ (MW 49.23 kDa) and P′′′(MW 47.23 kDa)) for the Fab, which represent different products from papain digestion of the mAb.[33] Shown in Figure 5.12b is a representative ESI mass spectrum acquired for the solution of Fab (5 μM) and carbohydrate library (1 μM each). Based on the reported affinities and the solution concentrations, approximately 1% of the Fab is bound to **L4** and 6% is bound to **L5**. Inspection of the mass spectrum reveals that the m/z of the -13 and -14 charge states of P′ are similar to those of the -13 and -14 charge states of P bound to **L4** or **L5**. These ions could, nevertheless, be partially resolved and quantified. In order to establish the presence of specific interactions between the Fab and both of the pentasaccharides, the $(Fab + L)^{-13}$ ions were isolated in the quadrupole mass filter and collisionally activated to release the bound ligands. As with the previous examples,

CID of the deprotonated complex leads to the facile loss of the deprotonated pentasaccharide ligands.

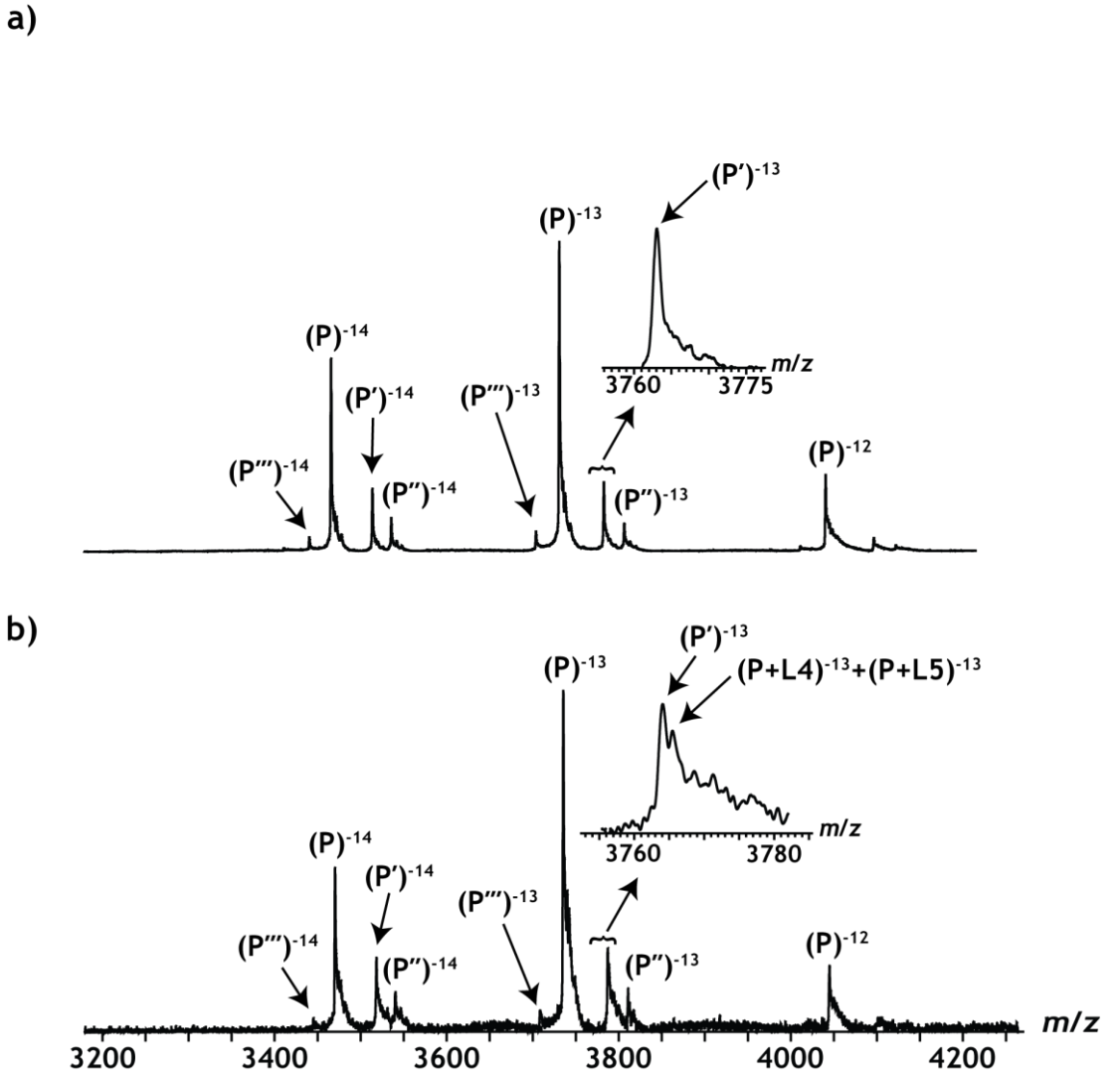


Figure 5.12 Representative ESI mass spectra acquired in negative ion mode for solutions of (a) Fab (5 μM) and (b) Fab (5 μM) and 203 carbohydrates (1 μM each). Four isoforms of the Fab were identified and labeled as P, P′, P′′ and P′′′.

Shown in Figure 5.13 are CID mass spectra measured for the released ligands (*post-release*) and for deprotonated **L4** and **L5** (*reference*). Overall, the *post-release* CID mass spectrum most closely resembles that of **L5**. However, clear differences in the relative abundance of the product ions are evident, suggestive of the presence of multiple carbohydrate structures. Shown in Figure 5.14 are the corresponding ATDs measured for the released ligands and for the individual **L4** and **L5** ions obtained directly from solution. Inspection of the ATD measured for the released ligands reveals two features (at 8.1 ms and 8.7 ms). This result, on its own, indicates that there are at least two distinct structures present. These features could, in principle, be due to the presence of two conformers of the same carbohydrate ion. However, comparison of the *post-release* ATD and those measured for the free, deprotonated **L4** and **L5** ions (*reference*) reveals that the two features observed for the *post-release* ATD are consistent with the presence of both ligands; the dominant feature at 8.7 ms in the ATD being consistent with the presence of **L5** and the minor feature at 8.1 ms reflecting a minor contribution from **L4**. Based on this analysis it is correctly concluded that both **L4** and **L5** bind to the Fab and that **L5** exhibits a significantly higher affinity than **L4**. Our laboratory is currently investigating whether the post-release ATDs can be quantitatively deconvoluted to establish the fraction of protein bound to individual isomeric ligands and, from that, the affinities for the individual ligands.

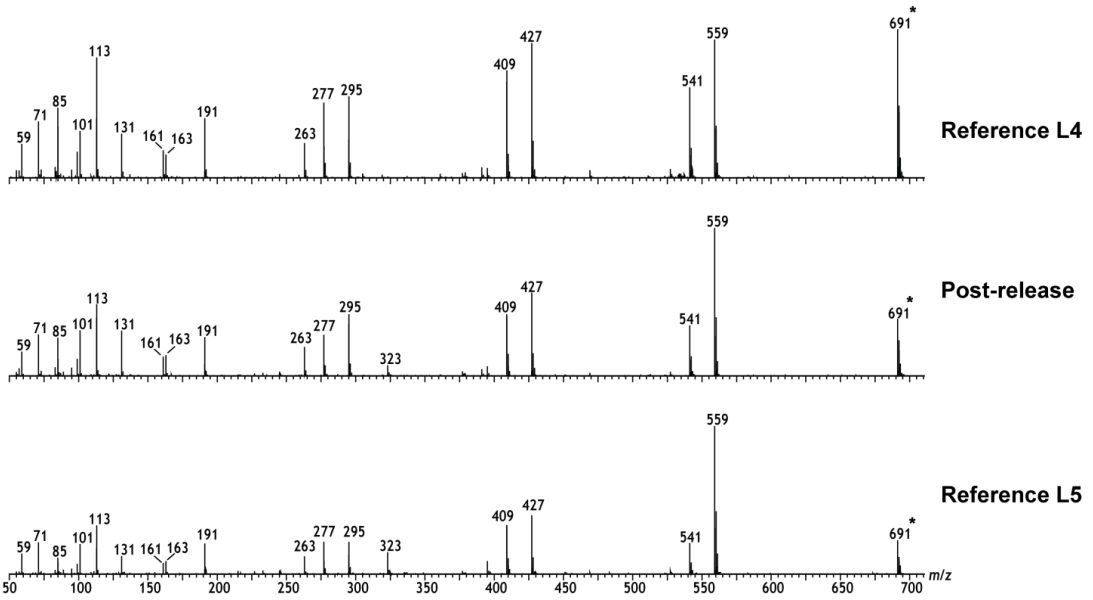


Figure 5.13 CID mass spectra measured at transfer voltage of 40V for deprotonated **L4** and **L5** ions obtained directly from solution (*reference*) and following their release from the $(Fab + L)^{-13}$ ions (*post-release*). The deprotonated **L4** and **L5** ions are denoted by *.

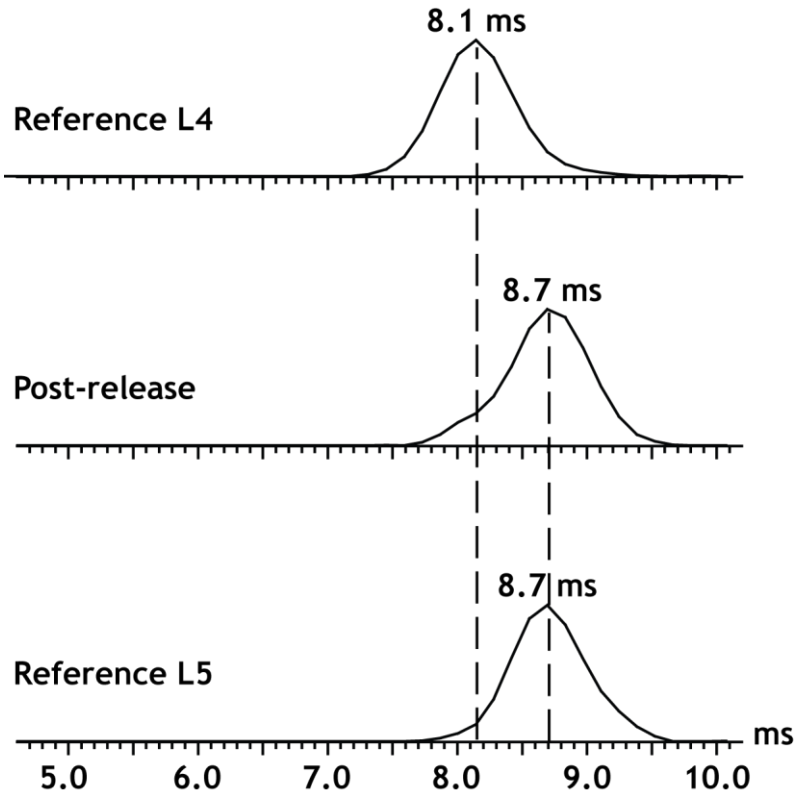


Figure 5.14 Arrival time distributions (ATDs) measured for the deprotonated ligands (L), **L4** and **L5**, following their release from the corresponding $(Fab + L)^{-13}$ ions (*post-release*). Also shown are the ATDs measured for the deprotonated **L4** and **L5** ions obtained directly from solution (*reference*).

5.4.3 Screening carbohydrate libraries for low affinity interactions

Given that many biologically important protein-carbohydrate interactions exhibit low affinities, ~10^3 M^{-1}, it was important to test the application of direct ESI-MS measurements for identifying multiple, low affinity interactions within a carbohydrate library. The *C*. *difficile* toxin B subfragment TcdB-B3 served as a model system for these measurements. Shown in Figure 5.15a is a representative ESI mass spectrum acquired in positive ion mode for a solution of TcdB-B3 (10 μM) and 23 carbohydrates (10 μM each). Because of the relatively high concentration of library components, αLA (15 μM) was added to the solution to serve as P_{ref} in order to correct mass spectra for nonspecific carbohydrate-protein binding. Analysis of the mass spectrum shown in Figure 5.15a reveals ion signals corresponding to the free TcdB-B3 and TcdB-B3 bound to nine different carbohydrates, **L6–L8**, **L10, L12**, **L36, L40**, **L47** and **L134**. However, also evident in the mass spectrum are low abundance ions corresponding to P_{ref} bound to seven of these carbohydrates, **L6–L8**, **L10, L12**, **L47** and **L134**. These results indicate the occurrence of nonspecific carbohydrate binding to the TcdB-B3 during the ESI process. Using the procedure described in the Experimental section, the mass spectrum was corrected for nonspecific binding. Shown in Figure 5.15b are the normalized distributions of carbohydrates bound to TcdB-B3 before and after correction for nonspecific binding. Using the corrected abundances, the corresponding K_a values were calculated and are shown in Table 5.3. Also listed in Table 5.3 are K_a values determined by ESI-MS of solutions of TcdB-B3 and the individual carbohydrate ligands. It can be seen that the K_a values determined by ESI-MS analysis of the library agree within a factor of three with those determined

from the individual ESI-MS measurements. These results indicate that direct ESI-MS screening of carbohydrate libraries can be used to identify low affinity protein-carbohydrate interactions and to estimate the corresponding association constants.

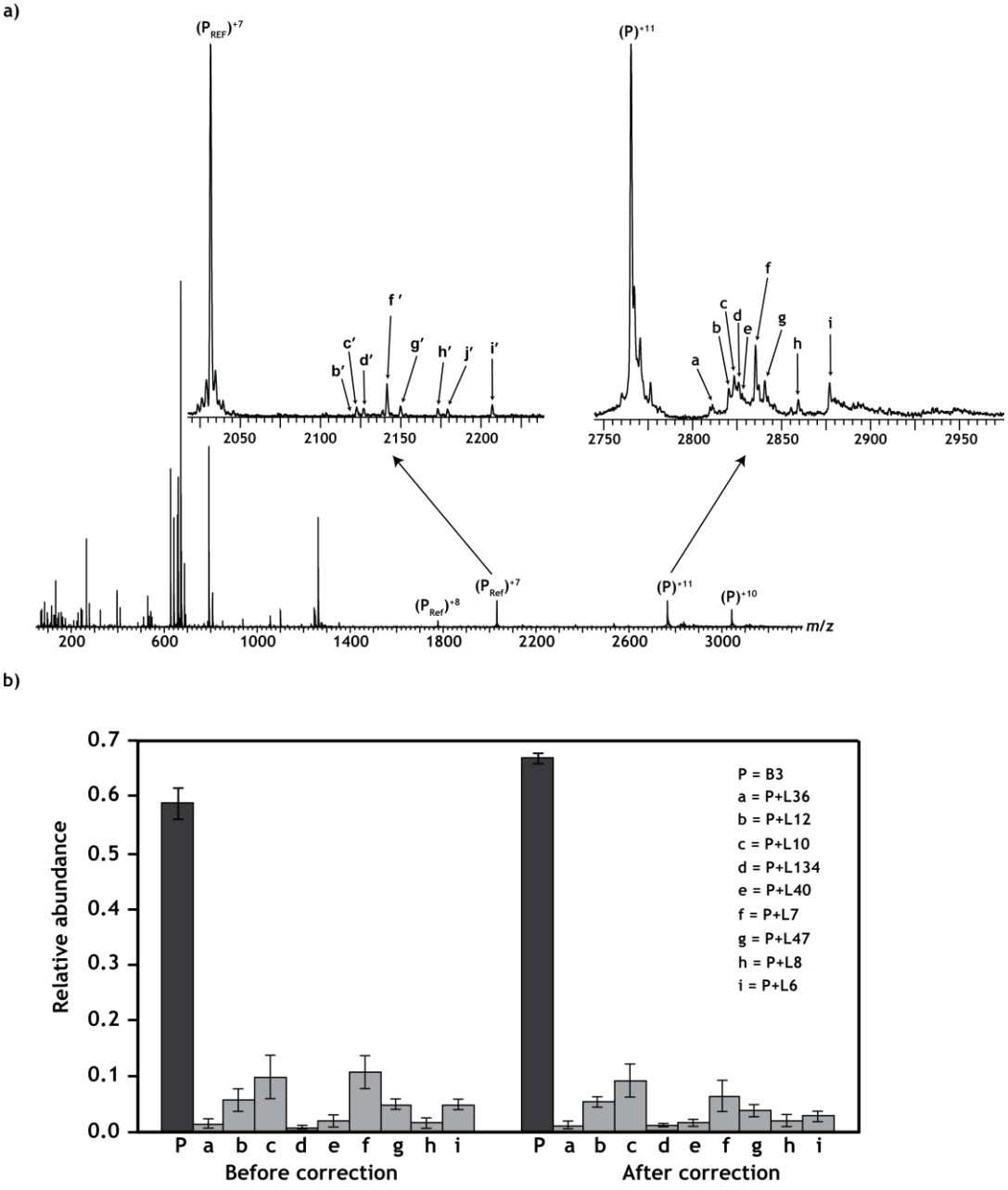


Figure 5.15 (a) Representative ESI mass spectrum obtained for a solution of TcdB-B3 (P, 10 μM), αLA (P_{ref}, 15 μM) and 23 carbohydrates (10 μM each)

acquired in positive ion mode, where **a** = (P+**L36**)$^{+11}$, **b** = (P+**L12**)$^{+11}$, **c** = (P+**L10**)$^{+11}$, **d** = (P+**L134**)$^{+11}$, **e** = (P+**L40**)$^{+11}$, **f** = (P+**L7**)$^{+11}$, **g** = (P+**L47**)$^{+11}$, h = (P+**L8**)$^{+11}$, **i** = (P+**L6**)$^{+11}$, while **b´** = (P_{ref}+**L12**)$^{+7}$, **c´** = (P_{ref}+**L10**)$^{+7}$, **d´** = (P_{ref}+**L134**)$^{+7}$, **f´** = (P_{ref}+**L7**)$^{+7}$, **g´** = (P_{ref}+**L47**)$^{+7}$, **h´** = (P_{ref}+**L8**)$^{+7}$, **i´** = (P_{ref}+**L6**)$^{+7}$, **j´** = (P_{ref}+**L137**)$^{+7}$. The measured and theoretical MWs of the complexes of TcdB-B3 complexes are: **L6**, measured 31648±1, theoretical 31648.4; **L7**, 31189±1, 31190.2; **L8**, 31452±2, 31454.6; **L10**, 31055±2, 31056.2; **L12**, 31024±1, 31025.7; **L36**, 30909±4, 30908.4; **L40**, 31123±2, 31126.8; **L47**, 31272±4, 31273.1; **L134**, 31087±2, 31087.7. (b) Normalized distributions of P bound to **L6–L8**, **L10, L12**, **L36, L40**, **L47** and **L134**, before and after correction for nonspecific binding using the reference protein method.

Table 5.3 Apparent association constants (K_a) for TcdB-B3 binding to **L6-L8**, **L10**, **L12**, **L36**, **L40**, **L47** and **L134** at 25 ºC and pH 7 determined by the direct ESI-MS assay performed on the individual ligands and a library consisting of 23 carbohydrates.[a]

Ligand	Apparent K_a (M^{-1}) *Library*	Apparent K_a (M^{-1}) *Individual*
L6	$(4.7 \pm 0.1) \times 10^3$	$(1.0 \pm 0.1) \times 10^4$
L7	$(9.8 \pm 0.2) \times 10^3$	$(1.1 \pm 0.4) \times 10^4$
L8	$(2.9 \pm 0.5) \times 10^3$	$(6.0 \pm 2.1) \times 10^3$
L10	$(1.5 \pm 0.3) \times 10^4$	$(1.1 \pm 0.2) \times 10^4$
L12	$(8.7 \pm 0.2) \times 10^3$	$(7.4 \pm 0.5) \times 10^3$
L36	$(1.7 \pm 0.1) \times 10^3$	$(2.4 \pm 1.2) \times 10^3$
L40	$(2.7 \pm 0.6) \times 10^3$	$(1.7 \pm 0.4) \times 10^3$
L47	$(5.8 \pm 0.3) \times 10^3$	$(1.9 \pm 0.5) \times 10^3$
L134	$(9.4 \pm 0.2) \times 10^2$	$(1.4 \pm 0.9) \times 10^3$

a. Errors correspond to one standard deviation.

5.5 Conclusions

In summary, the first detailed study of the applications of a CaR-ESI-MS assay for carbohydrate library screening against carbohydrate-binding proteins is described. Direct ESI-MS measurements were performed on solutions containing a target protein and a library of carbohydrates containing multiple, specific ligands with affinities in the 10^3 to 10^6 M^{-1} range. Ligands with moderate affinity were successfully detected from mixtures containing >200 carbohydrates at concentrations as low as 0.25 μM each. Additionally, the absolute affinities were estimated from the abundance of free and ligand bound protein ions. Multiple, low affinity ligands were also successfully detected in a small library of 23 carbohydrates. However, identification of the specific interactions required the use of a reference protein to correct the mass spectra for the occurrence of nonspecific carbohydrate-protein binding during the ESI process. The release of the carbohydrate ligands in their deprotonated form using CID performed on the deprotonated protein-carbohydrate complexes was also successfully demonstrated. Comparison of the arrival time distributions measured for the deprotonated carbohydrate ions, following their release, allowed for the positive identification of isomeric ligands.

5.6 Literature cited

1. Roldos, V.; Javier, C. F.; Jimenez-Barbero J. *ChemBioChem* **2011**, *12*, 990-1005.
2. 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.
3. Avci, F. Y.; Kasper, D. L. *Annu*. *Rev*. *Immunol*. **2010**, *28*, 107 –130.
4. Holgersson, J.; Gustafsson, A.; Breimer, M. E. *Immunol*. *Cell Biol*. **2005**, *83*, 694–708
5. Cipolla, L.; Araujo, A. C.; Bini, D.; Luca, G. ;Laura, R. ; Nasrin, S. *Expert Opin*. *Drug Discov*. **2010,** *5*, 721-737.
6. Lundquist, J. J.; Toone, E. J. *Chem. Rev.* **2002**, *102*, 555-578.
7. A. Bundle, D. R. Methods Enzymol. **1994**, *247*, 288-305. B. Bundle, D. R. In Carbohydrates Hecht, S., Ed.; Oxford University Press: Oxford, **1998**; pp 370-440.
8. Zhang, B. Y.; Chan, N.; Schriemer, D. C.; Palcic, M. M.; Hindsgaul, O. *Glycobiology* **1999**, *9*, 14.
9. Larsen, K.; Thygesen, M. B.; Guillaumie, F.; Willats, W. G. T.; Jensen, K. J. *Carbohydrate Res*. **2006**, *341*, 1209-1234.
10. Disney, M. D.; Peter H. Seeberger, P. H. *Drug Discov*. *Today: TARGETS* **2004**, *3*, 151-158.
11. Fais, M.; Karamanska, R.; Russell, D. A.; Field, R. A. *J*. *Cereal Sci*. **2009,** *50*, 306–311.

12. Loo, J. A. *Mass Spectrom*. *Rev*. **1997,** *16*, 1–23.

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

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

15. Heck A. J. R.; Van Den Heuvel R. H. H. *Mass Spectrom. Rev*. **2004**, *23*, 368-389.

16. Gabelica, V.; Rosu, F.; De Pauw, E. *Anal*. *Chem*. **2009**, *81*, 16, 6708-6715.

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

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

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

20. Rademacher, C.; Shoemaker, G. K.; Kim, H. S.; Zheng, R. B.; Taha, H.; Liu, C.; Nacario, R. C.; Schriemer, D. C.; Klassen, J. S.; Peters T.; Lowary, T. L. *J. Am. Chem. Soc.* **2007**, *129*, 10489-10502.

21. Jorgensen, T. J. D.; Roepstorff, P.; Heck, A. J. R. *Anal*. *Chem*. **1998**, *70*, 4427-4432.

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

23. Yu, Y. H.; Kirkup, C. E.; Pi, N.; Leary, J. A. *J*. *Am*. *Soc*. *Mass Spectrom*. **2004**, *15*, 1400-1407.

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

25. Sun, J.; Kitova, E. N.; Klassen, J. S. *Anal*. *Chem*. **2007**, *79*, 416-425.

26. Bagal, D.; Kitova, E. N.; Liu, L.; El-Hawiet, A.; Schnier, P. D.; Klassen, J. S. *Anal*. *Chem*. **2009**, *81*, 7801-7806.

27. Abzalimov, R. R.; Dubin, P. L.; Kaltashov, I. A. *Anal*. *Chem*. **2007**, *79*, 6055-6063.

28. Cheng, X.; Chen, R.; Bruce, J. E.; Schwartz, B.; Landerson, G. A.; Hofstadler, S. A.; Gale, D. C.; Smith, R. D. *J*. *Am*. *Chem*. *Soc*. **1995**, *117*, 8859-8860.

29. Gao, J.; Cheng, X.; Chen, R.; Sigal, G. B.; Bruce, J. E.; Schwartz, B. L.; Hofstadler, S. A.; Anderson, G. A.; Smith, R. D.; Whitesides, G. M. *J*. *Med*. *Chem*. **1996,** *39*, 1949-1955

30. Wigger, M.; Eyler, J. R. ; Benner, S. A.; Li, W.; Marshall, A. G. *J*. *Am*. *Soc*. *Mass Spectrom*. **2002**, *13*, 1162–1169

31. Cederkvist, F. H.; Zamfir, A. D.; Bahrke, S.; Eijsink, V. G. H.; Sørlie, M.; Jasna Peter-Katalinic´, J.; Peter, M. G. *Angew*. *Chem*. *Int*. *Ed*. **2006**, *45*, 2429 – 2434.

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

33. Rademacher, C.; Shoemaker, G. K.; Kim, H. S.; Zheng, R. B.; Taha, H.; Liu, C.; Nacario, R. C.; Schriemer, D. C.; Klassen, J. S.; Peters T.; Lowary, T.L. *J. Am. Chem. Soc.* **2007**, *129*, 10489-10502.

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

35. Greco, A.; Ho, J. G. S.; Lin, S. J.; Pacilic, M. M.; Rupnik, M.; Ng, K. K. S. *Nat. Struct. Mol. Biol.* **2006**, 13, 460-461.

36. Ayers, J. D.; Lowary, T. L.; Morehouse, C. B.; Besra, G. S. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 437–442.

37. D'Souza, F. W.; Ayers, J. D.; McCarren, P. R.; Lowary, T. L. *J. Am. Chem. Soc.* **2000**, *122*, 1251–1260.

38. Yin, H.; D'Souza, F. W.; Lowary, T. L. *J. Org. Chem.* **2002**, *67*, 892–903.

39. Joe, M.; Bai, Y.; Nacario, R. C.; Lowary, T. L. *J. Am. Chem. Soc.* **2007**, *129*, 9885–9901.

40. Rademacher, C.; Shoemaker, G. K.; Kim, H. S.; Zheng, R. B.; Taha, H.; Liu, C.; Nacario, R. C.; Schriemer, D. C.; Klassen, J. S.; Peters, T.; Lowary, T. L. *J. Am. Chem. Soc.* **2007**, *129*, 10489–10502.

41. Completo, G. C.; Lowary, T. L. *J. Org. Chem.* **2008**, *73*, 4513–4525.

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

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

44. 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*, 5172-5182.

45. a) http://webbook.nist.gov/chemistry/ b) Jebber, K. A.; Zhang, K.; Cassady, C. J.; Chung-Phillips, A. *J. Am. Chem. Soc.* **1996**, *118*, 10515-10524. c) Cassady, C. J. *Gas-Phase Basicities of Small Carbohydrates*; The 41st ASMS Conference on Mass Spectrometry, San Francisco, CA, May 31, **1993**.

46. Salpin, J. Y.; Tortajada, J. *J*. *Mass Spectrom*. **2004**, *39*, 930-941.

47. a. In *Gas-Phase Ion chemistry*; Bowers, M. T., Ed.; Academic Press, New York, **1979**, Vol. 2, Ch. 11. b. Cumming, J. B.; Kebarle, P. *Can*. *J*. *Chem*. **1978**, *56*, 1-9.

48. Domon, B.; Costello, C. E. *Glycoconjugate J*. **1988**, *5*, 397-409.

49. Liu, Y.; Clemmer, D. E. *Anal*. *Chem*. **1997**, *69*, 2504-2509.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

This work describes the development and application of ESI-MS methods to study non-covalent protein-carbohydrate interactions. In Chapter 2, we used the direct ESI-MS assay to quantify the affinities of two *C. difficile* toxins subfragments (TcdA-A2 and TcdB-B1) with a library of 21 human milk oligosaccharides (HMOs) representing the most abundant acidic and neutral sugars in human milk.[1] A significant finding of this study is that both toxins subfragments bind, albeit weakly, to many of the HMOs tested, where five of the studied HMOs bind both TcdA-A2 and TcdB-B1. We explored the binding modalities of HMOs with *C. difficile* toxin fragments through molecular docking simulations. The molecular docking results suggest that, despite a number of differences in amino acid sequence between TcdA and TcdB, the general mode of carbohydrate recognition may be conserved and that lactose disaccharide appears to occupy the central portion of the carbohydrate binding site for both toxins. To investigate the inhibitory potential of HMOs on TcdA and TcdB, Verocytotoxicity neutralization assays were performed using six fractions of HMOs extracted from human milk samples. The results of Verocytotoxicity neutralization assay reveal that the HMOs do not significantly inhibit the cytotoxic effects of TcdA or TcdB, which is attributed to the weak intrinsic affinities that the toxins exhibit towards the HMOs.

In Chapter 3, a new ESI-MS binding assay, called the *reference ligand* ESI-MS method, was described. This method combines the direct ESI-MS assay and

competitive ligand binding to quantify the binding affinities of protein-ligand interactions that undergoes in-source dissociation.[2,3] The *reference ligand* ESI-MS method relies on the introduction of a reference ligand (L_{ref}) which binds to the same binding site as the ligand of interest with a known binding affinity and forms a stable gas phase complex. Quantifying the protein-ligand interactions that are unstable in the gas phase is accomplished through measuring the relative abundance of L_{ref}-bound protein to free protein, which is sensitive to the presence of other ligands in solution that compete for the same binding site, by the direct ESI-MS assay. The reliability of the *reference ligand* ESI-MS method was demonstrated using two carbohydrate-binding proteins, Se155-4 scFv and ConA, with their specific monosaccharide ligands. The association constants obtained with this new method are in a good agreement with the ITC-derived values.[4,5]

In Chapter 4, the *proxy protein* ESI-MS method was described. This new method allows the quantification of the protein-ligand complexes that cannot be directly detected by ESI-MS either due to the microheterogeneity of the protein or due to its molecular weight that exceeds the upper mass limit of the mass analyzer used. The *proxy protein* ESI-MS method involves using a *proxy protein* that binds to the ligand of interest with a known binding constant and its bound and free forms can be readily detected by direct ESI-MS assay. The homotrimeric tailspike (TSP) protein of the bacteriophage P22[6], as well as an endorhamnosidase point mutant, and its octa- and dodecasaccharide ligands which comprise two and three O-antigen repeats from *Salmonella* Typhimurium,[7] were used as model systems in this study. The binding constants quantified at 10 and 25 °C agree to a great extent to the reported

fluorescence quenching assay data.[6] In addition, we developed a modified form of the assay, which accounts for the real-time changes in concentration of the ligand due to the hydrolyzing activity of the wild type TSP toward the dodecasaccharide ligand. This allows for the first time to quantify the K_a value for wild type TSP binding to the dodecasaccharide ligand.

In Chapter 5, we carried out a detailed investigation into the use of the catch and release (CaR) ESI-MS approach for screening carbohydrate libraries against target proteins for specific interactions. In this study, three protein systems were studied together with their specific carbohydrate ligands. We demonstrated that in a single experiment, it is possible to detect the complexes of Se155-4 scFv with three specific ligands in a library of 204 carbohydrates. Moreover, the absolute affinities were estimated from the abundance of free and ligand-bound protein ions determined directly from the ESI mass spectrum and they agreed to a great extent with the values determined from individual measurements. We also demonstrated the applicability of the CaR-ESI-MS in cases where the target protein binds to isomeric ligands. We screened CS35 Fab against a library of 203 carbohydrates including two specific isomeric ligands. The use of ion mobility separation allowed for the positive identification of isomeric ligands through comparing the ATDs of the released ligands with references. We also successfully screened a library of 23 carbohydrates against subfragment B3 of *C. difficile* toxin B and identified nine low affinity specific ligands.

6.2 Future work

Over the last two decades ESI-MS has emerged as an important addition to the arsenal of tools for the identification and quantification of protein-ligand interactions *in vitro*. Despite the successful implementation of ESI-MS in studying many protein non-covalent complexes, there are still areas for improvements to enhance the versatility and utility of the technique and to tackle the challenges that limit its widespread adoption.

A relevant future work to chapter 4 is to use ESI-MS to quantify ligand binding to very large protein complexes, like whole viruses and their particles, in cases where the *proxy protein* method cannot be applied. For example when no suitable proxy protein is available or the target protein could not be prepared in sufficient concentrations to be used to compete for the ligand with the proxy protein or when the presence of the large target protein affects the response factor of the proxy protein and its complexes. A possible method to deal with these situations depends on monitoring the change in intensity of the free ligand upon changing the concentration of the target protein. Wortmann et al. has reported a similar strategy which relies on the use of a reference ligand (L_{ref}), with known affinity for the protein of interest, and the determination of the relative abundance of the ligand and L_{ref} by ESI-MS.[8] A related method was used to establish the relative ligand affinities by measuring the signal intensity ratio of the ligands upon increasing the target protein concentration.[9]

The proposed approach relies on using an internal standard (IS) with very similar ionization efficiency as the ligand of interest (L). The experiment would involve monitoring the change in the intensity ratio of the IS to L (I_{IS}/I_L) upon titrating

equimolar mixtures of L and IS with the target protein. The equilibrium concentration of L can be determined using the calibration curve of the dependence of I_{IS}/I_L on changing L concentrations in the absence of the target protein. This method benefits from better sensitivity and better resolution for small molecule detection in the lower mass range which allows the use of very low concentrations of the ligands and the target protein. An important point to be investigated, in this approach, is the effect of the presence of these large target proteins on the relative response factor of L and IS.

As described in chapter 5, the CaR-ESI-MS can be used to screen proteins against libraries of carbohydrates where the identity of the released ligands is confirmed through comparing the ATDs of the released ligands after the ion mobility separation with that of reference ligands. A clear extension to this project is to develop methods to quantify, or at least rank, the binding affinities of the isomeric ligands. A possible way to achieve this goal is to use the ATDs peak areas of the released ligands. The following section describes the quantification of the association constants of isomeric ligands using this approach.

For a solution of protein (P) and two isomeric ligands, (L_1 and L_2), the ratio of the ATDs areas (R^t) of the released ligands is assumed to be proportional to the ratio of the equilibrium concentrations of the corresponding complexes ($[PL_1]$ and $[PL_2]$):

$$R^t = \frac{[PL_2]}{[PL_1]} \qquad (6.1)$$

The abundance ratio (R) of ligand-bound and unbound protein ions, obtained from the mass spectrum, is expected to be equivalent to the equilibrium concentration ratio in solution according to eq. 6.2

$$R = \frac{[PL_1]+[PL_2]}{[P]} \tag{6.2}$$

Combining eq. 6.1 and 6.2, gives

$$[P] = \frac{[PL_1](R^t+1)}{R} \tag{6.3}$$

Using the mass balance equation for the initial protein concentration $[P]_o$ (eq. 6.4) and eq. 6.3, the equilibrium concentration of $[PL_1]$ could be calculated using eq.6.5:

$$[P]_o = [P]+[PL_1]+[PL_2] \tag{6.4}$$

$$[PL_1] = \frac{[P]_o R}{(R^t+1)(R+1)} \tag{6.5}$$

Using the mass balance equation for the initial L_1 concentration $[L_1]_o$ (eq. 6.6) and eq. 6.5, the equilibrium concentration of $[L_1]$ is calculated according to eq. 6.7:

$$[L_1]_o = [L_1]+[PL_1] \tag{6.6}$$

$$[L_1] = [L_1]_o - \frac{[P]_o R}{(R^t+1)(R+1)} \tag{6.7}$$

Using equations 6.3, 6.5 and 6.7, the binding constant of L_1 ($K_{a,L1}$) is equal to eq. 6.8:

$$K_{a,L1} = \frac{R(1+R)}{[L1]_o(R^t+1)(R+1)-[P]_o R} \tag{6.8}$$

Following the same procedure, the binding affinity of L_2 could be quantified.

The proposed approach has been tested on two protein systems. The first system is Se155-4 scFv and two isomeric ligands (α-D-Tal-(1→2)-[α-D-Abe-(1→3)]-α-D-Man-OCH_3 and α-D-Glc-(1→2)-[α-D-Abe-(1→3)]-α-D-Man-OCH_3). The second system is CS35 Fab and two isomeric ligands (β-D-Ara-(1→2)-α-D-Ara-(1→5)-[α-D-Ara-(1→3)]-α-D-Ara-(1→5)-α-D-Ara-OCH_3 and β-D-Ara-(1→2)-α-D-Ara-(1→3)-[α-D-Ara-(1→5)]-α-D-Ara-(1→5)-α-D-Ara-OCH_3). The preliminary results indicate that the binding affinities obtained using this approach are within a factor of 2-3 from those obtained from the individual measurements. Further research is required to enhance the resolutions of the ATDs of the released ligands especially for epimers by studying, for example, the effect of using metal cations to induce the formation of the corresponding metal adducts of the released ligands. Also a related possible point is to explore using the intensity of the characteristic fragments from the fragmentation spectrum of the released ligands to estimate the corresponding binding constants.

Another logical extension for the CaR-ESI-MS assay is to screen natural extracts. This approach could be applied to natural extracts after minor separation steps as the CaR-ESI-MS assay, in principle, can screen libraries of several thousands of compounds simultaneously. Among the many possible extracts (plant, human and microbial), the human milk is a very rich source of diverse carbohydrates. The human milk oligosaccharides (HMOs) are the third largest solid component (0.5-1%) of the human milk.[1] HMOs have proven protective activities against many pathogens and their toxins. This is believed to be, at least in a part, due to the structural similarity between HMOs and the natural receptors of these pathogens which divert the pathogens from binding to their native receptors on the cell surfaces.[10] The application

of CaR-ESI-MS assay in screening HMOs fractions will allow the rapid screening of this potential source of carbohydrate ligands against microbial pathogens (including viral and bacterial fragments) and their toxins like Norovirus and heat stable *E.coli* toxins. Research efforts are required to explore different purification protocols to minimize the salt content of the HMOs extracts, which complicates the mass spectra due to the formation of metal adducts especially (Na^+ and K^+). Another important point of research in this project is to identify and differentiate isomeric ligands, which are very common in HMOs, through developing reference databases for the fragmentation patterns and ATDs for the most abundant isomers.

6.3 Literature cited

1. Bode, L. *J*. *Nutr*. **2006** *136*, 2127–2130.
2. Van Dongen, W. D.; Heck A. J. R. *Analyst* **2000**, *125*, 583-589.
3. Sun, J.; Kitova, E. N.; Klassen, J. S. *Anal*. *Chem*. **2007**, *79*, 416-425.
4. 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*, 5172-5182.
5. Castro, S.; Duff, M.; Snyder, N. L.; Morton, M.; Kumar, C. V.; Peezuh, M. W. *Org*. *Biomol*. *Chem*. **2005**, *3*, 3869-3872.
6. Baxa, U.; Steinbacher, S.; Miller, S.; Weintraub, A.; Huber, R.; Seckler, R. *Biophys J*. **1996**, *71*, 2040–2048.
7. Sifurskjold, B. W.; Altman, E.; Bundle, D. R. *Eur*. *J*. *Biochem*. **1991**, *197*, 239-246.

8. Jecklin, M. C.; Touboul, D.; Jain, R.; Toole, E. N.; Tallarico, J.; Drueckes, P.; Ramage, P.; Zenobi, R. *Anal*. *Chem*. **2009,** *81*, 408–419.

9. Wortmann, A.; Jecklin, M. C.; Touboul, D.; Badertscher, M.; Zenobi, R. *J*. *Mass Spectrom*. **2008**, *43*, 600-608.

10. Newburg, D. S. *J*. *Anim Sci*. **2009**, *87*, 26-34.