MASS SPECTROMETRY-BASED BIOMOLECULAR RECOGNITION AND RESPONSE FACTOR INVESTIGATIONS USING ELECTROSPRAY IONIZATION

by

MISJUDEEN RAJI

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

August 2008

Dedication

To my loving parents; every height I reach, and every honor I earn, is by dint of Allah's mercy, and your love, care, and Dua'.

ACKNOWLEDGEMENTS

All praise is due to Allah (GOD), the Lord of incomparable Majesty. I give him all the glory for seeing me through this dissertation work successfully. One could say my story is a good example of the saying “If God takes you to it, He will surely see you through it”.

My profound gratitude goes to my parents, my brothers and sisters, and my loving wife, for their unwavering support throughout the period of my study. Verily, without their encouragement and backing, the journey thus far would have been much more arduous. Next I will like to thank the man who has left perhaps the most indelible marks on my heart throughout my sojourn in graduate school, Dr Kevin Schug. His immense support, advice, encouragement, reassurance, and occasional push have been highly invaluable to the success of my work. From him, I have not only learned mass spectrometry and analytical chemistry, but I have also learned good leadership.

I also want to express my appreciation to my dissertation committee members, Dr Richard Timmons, and Dr Richard Guan. Their advice and encouragement throughout my dissertation work have been more than helpful. I also thank Dr Daniel Armstrong, Dr Jung-Mo Ahn, Dr Seoung Kim, Miss Bao Ye, and Miss Chivalai Temiyasathit for their advice and collaboration on some aspects of my dissertation work.

I also thank my colleagues in the Schug’s lab for their support and cooperation throughout my tenure in the Schug’s lab. I am very grateful to Dr Petr Frycak, who introduced me to computer programming for data analysis.

Finally, I want to thank Dr (Mrs.) Mary Kinsel and Dr Zaneer Segu, whose words of encouragement at the time I was finishing my Masters degree went a long way in my decision to pursue Doctoral degree. And to all friends and well wishers, I say “Thank you all for being there”.

July 15, 2008

ABSTRACT

MASS SPECTROMETRY-BASED BIOMOLECULAR RECOGNITION AND RESPONSE FACTOR INVESTIGATIONS USING ELECTROSPRAY IONIZATION

Misjudeen Raji, PhD.

The University of Texas at Arlington, 2008

Supervising Professor: Kevin Schug

Integrins are transmembrane proteins found in almost all cells. They play vital roles in many biological processes, such as angiogenesis and signal transduction, through their interactions with different extracellular matrix ligands. These interactions are fostered by forming noncovalent bonds with the ligands. Integrin involvement in angiogenic cancer metastasis has sparked the interest of many researchers in the field of oncology. Many cutting edge research works are now focused on studying the noncovalent interactions of integrins with potential drug compounds formulated from peptide mimics and bearing the putative RGD amino acid sequence through which integrins are known to bind their ligands. Most of the published works on the study of noncovalent interactions of integrins with their ligands have been carried out using traditional techniques such as X-Ray crystallography, NMR, and immunoassay development. The introduction of ESI-MS to the list of viable techniques for probing the integrin-RGD noncovalent interactions is a logical and much needed approach due to its versatility and high throughput capability compared to the more traditional techniques.

The experiments described in this dissertation serve to establish, for the first time, the amenability of ESI-MS for assessing integrin-RGD binding affinity using only peptide fragments that represent the binding region of the intact integrin protein molecule. Using information from crystallographic data obtained from the literature about the amino acid sequence of the integrin

binding domain, peptide fragments were synthesized having these amino acid sequences and tested for affinity to different short peptide ligands. Important considerations regarding analyte response factors during ESI-MS process, and how these impact binding affinity determination in ESI-MS, were also addressed experimentally and in conjunction with statistical tools.

Three major challenges are noted with this new approach. The first has to do with marked difference between the conformations of the peptide fragments employed compared to that of the tertiary structure of the integrin molecule. Also, the binding constants in these experimental conditions may differ from actual physiological conditions. And lastly is the question of how the response factors of host, guest, and complex affect the measured binding affinities.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1

DISSERTATION OVERVIEW

"Research is to see what everybody else sees and to think what no one else has thought"

- Albert Szent-Gyorgyi

1.1 Introduction

Noncovalent interactions in biological systems are crucial processes and are vital to the proper functioning of many biomolecules. Many biological activities such as cell-cell signaling, cell migration and cell adhesion are mediated by the incessant formation and disruption of noncovalent forces between large biomolecules (such proteins) and their ligands. Another aspect of the role of noncovalent interactions in the biological systems is in 'targeted' drug delivery, where a particular biomolecule that is known to be involved in a disease-related biological activity is targeted for drug delivery by taking advantage of its specific interaction with known ligands; such ligands are then employed as drug carriers for the particular disease in which the target biomolecule is involved.

For this work, the integrin protein molecule has been chosen for study. Integrins are transmembrane proteins found in almost all cells. They are known to mediate several biological activities in the cell via the formation of noncovalent complexes with extracellular matrix proteins. Particularly, they are involved in angiogenesis, which is the process by which cells proliferate through the formation of new blood vessels. Due to the vital role of angiogenesis in cancer metastasis, integrins are reputed to be an important class of proteins aptly suitable as targets for peptidomimetic drug delivery. A number of landmark studies have been carried out to elucidate the roles of integrins in cancer metastasis as well as to identify potential peptide-based ligands that interact with integrins (mainly those containing the Arg-Gly-Asp (RGD) amino acid sequence). These studies have mostly been carried out using traditional solution phase

techniques to study the noncovalent interactions between integrins and these ligands. Many of these techniques lack specificity and high throughput capability. In order to keep up with the high paced developments in the drug discovery research endeavors, there is need for a technique to fill the vacuum that is not covered by the traditional techniques. Mass spectrometry is one such technique that has the requisite figures of merit to fulfill this need.

The focus of this work is to investigate the amenability of ESI-MS for evaluating the solution phase binding interactions between integrin fragments and RGD peptides, with the goal of establishing some fundamentals of ESI-MS application for this particular biological system as well as aiding the development of rapid and high throughput screening methods for synthetic peptidomimetic drug molecules. The integrin fragments employed for this study were synthesized to mimic the putative binding pockets of the whole integrin molecules using information obtained from earlier crystallographic studies carried out on these proteins as reported in the literature.

This dissertation comprises two central themes. The first deals with the determination of solution phase binding interactions of several integrin fragments with different RGD peptides using a combination of techniques. This work demonstrates, for the first time, the applicability of ESI-MS for characterizing integrin-RGD binding properties in such a fashion that makes possible a high-throughput screening of different potential ligands by using integrin fragments comprising the amino acid sequence of the protein binding region. Data obtained from these ESI-MS studies show some correlation with those from capillary electrophoresis (CE) measurements. These findings further corroborate reports in the literature pertaining to the potential of ESI-MS as a viable alternative for the determination of solution phase binding constants of host-guest complexes. It is expected that this will pave the way for further research in this area, especially given the wide variety of permutations possible with the amino acid sequence of the RGD peptides and the host of combinations possible with several integrins available.

The second aspect pertains to the study of the effects of both analyte physicochemical parameters and instrument parameters on analyte response in ESI-MS using a set of model

peptides as analytes and applying multivariate statistical tools for data treatment. This part of the dissertation highlights the possibility of employing statistical tools for delineating the parameters that are most relevant to response factor determination in ESI-MS measurements of small molecules. The motivation behind this aspect of the work comes from the need to better understand how an analyte's properties influence its ESI-MS response, and how such knowledge can be used to predict the response factor of the analyte. This information will greatly facilitate the process of determining host-guest binding affinities in particular (where complex response factor has hitherto been considered similar to the host response factor to simplify calculations), and analyte quantitation in general. Multivariate statistical tools such as multi-linear regression were used to model the effects of analytes physicochemical parameters while full factorial chemometrics analysis was used to investigate the effects of instrumental parameters.

CHAPTER 2

GENERAL BACKGROUND

2.1 Introduction

Molecular recognition is a process by which two or more molecules interact with one another through the formation of noncovalent bonds. It is an important aspect in chemical research that is currently receiving wide attention for various applications. Molecular recognition is being employed for highly selective separations as well as in the design of chemical sensors. A recent application that has spurred great interest is in the area of drug discovery [1]. This stems from the fact that molecular recognition permits the fast screening of potential drug candidates, as well as rapid identification of cellular targets of drug molecules.

Molecular recognition events in biological systems are of paramount importance because they mediate many cellular activities. For example, the binding of a hormone (e.g. insulin) to its trans-membrane receptor protein can lead to a cascade of events inside the cell; which in turn can result in the secretion of an enzyme, the activation of a metabolic pathway, or other biological phenomena. Many diseases are also known to result from the disruption of the noncovalent binding interactions between a receptor and its target molecule. Another example of molecular recognition in the biological system is the binding of integrins (a class of transmembrane protein molecules) to extracellular matrix proteins. This binding phenomenon is known to mediate several cellular processes like angiogenesis and apoptosis which are important processes in tumor development and cancer metastasis. It is therefore very important to understand the mechanism of these binding events, not only from the standpoint of molecular biology, but also in designing target-specific drugs to combat diseases.

Many analytical tools have been employed to study these binding mechanisms. Of them all, mass spectrometry stands out because of its high speed, sensitivity and versatility. Mass

spectrometry provides a very powerful technique for studying proteins and protein-ligand complexes. One obvious advantage is its ability to measure the molecular weights of protein molecules in their bound and unbound forms. Mass spectrometry techniques that have been employed for studying noncovalent complexes include electrospray ionization-mass spectrometry (ESI-MS) and matrix-assisted laser desorption/ionization (MALDI). Of these two, ESI-MS is the preferred method because, being a "softer" ionization technique, it can be tuned to allow the introduction of noncovalent complexes into the instrument directly from solution without disrupting binding. The earliest example demonstrating the applicability of ESI-MS for studying noncovalent protein complexes is in the works of Ganem et al [2] who studied the binding of FKBP (an immunosuppressive protein) to FK506 and rapamycin (immunosuppressive agents). From then onwards, a tremendous amount of research efforts have gone into the study of noncovalent interactions using ESI-MS. A host of review articles on this application of ESI-MS have been published in the literature [3-8]. Both solution phase (e.g. host-guest complex screening, titration, and competitive binding) and gas phase (e.g., kinetic method and collision induced dissociation) techniques are typically employed when using ESI-MS for molecular recognition studies.

2.2 Molecular Recognition: The Case of Integrin and RGD Peptides.

The phenomenon of molecular recognition can be simply defined as the formation of multi-molecular complexes when two or more molecules that possess complementary characteristics in their functional and structural nature bind together. These complexes are usually formed as a result of noncovalent interactions taking place at the binding sites of the participating molecules. The prevalent forces responsible for noncovalent interactions in molecular recognition events are hydrogen bonding, ionic, hydrophobic (or solvophobic), ion-dipole, and dipole-dipole interactions. Some of these forces are described in Table 2.1

Table 2.1 Interaction names and potentials for specific non-covalent interactions

Interaction Type	Formula
Charge-charge (Coulomb energy)	$\frac{Q_1 Q_2}{4\pi\varepsilon\varepsilon_0 r}$
Charge-dipole (fixed dipole)	$-\frac{Qu\cos\theta}{4\pi\varepsilon\varepsilon_0 r^2}$
Charge-dipole (freely rotating dipole)	$-\frac{Q^2u^2}{6(4\pi\varepsilon\varepsilon_0)^2 kTr^4}$
Dipole-dipole (fixed dipole)	$-\frac{u_1 u_2}{4\pi\varepsilon\varepsilon_0 r^3}(2\cos\theta_1\cos\theta_2-\sin\theta_1\cos\phi\sin\theta_2)$
Dipole-dipole (freely rotating pole) (Keesom energy)	$-\frac{u_1^2 u_2^2}{3(4\pi\varepsilon\varepsilon_0)kTr^6}$
Charge-nonpolar	$-\frac{Q^2\alpha}{2(4\pi\varepsilon\varepsilon_0)^2 r^4}$
Dipole-nonpolar (fixed dipole)	$-\frac{u^2\alpha(1+3\cos^2\theta)}{2(4\pi\varepsilon\varepsilon_0)^2 r^6}$
Dipole-nonpolar (freely rotating pole) (Debye energy)	$-\frac{u^2\alpha}{(4\pi\varepsilon\varepsilon_0)^2 r^6}$
Nonpolar-nonpolar (London Dispersion energy)	$-\frac{3}{2}\frac{\alpha_{01}\alpha_{02}}{(4\pi\varepsilon\varepsilon_0)^2 r^6}\frac{I_1 I_2}{I_1+I_2}$
Hydrogen bond	Special, directed interaction
Hydrophilic	Special interaction
Hydrophobic	Special interaction

Q = charge, r = distance, α = polarizability, ε = dielectric constant, I = first ionization potential, θ = angle between dipole and vector connecting the interacting particles, ϕ = polar angle of second dipole relative to the first. (Adapted from reference [3].

In biological systems, noncovalent interactions are very important because they mediate biological activities and cellular functions. For example, the disruption of noncovalent interactions in a complex that is involved in a normal cellular process may lead to a diseased state. Hence, a critical first step towards understanding some of the disease processes is to fully comprehend the structures and mechanisms of these noncovalent complexes [9]. This understanding will help in designing target-specific drug compounds capable of combating the disease and restoring normal cellular functions.

Another example to demonstrate the importance of molecular recognition in biological systems is the transport of biologically important molecules (e.g. DNA, hormones) across the cell membrane. Many trans-membrane proteins are known to be involved in the transport of molecules from the intracellular environment to the extra-cellular matrix (or vice versa) by simply binding to and facilitating transport of their target molecules [10]. Other cellular activities like cell-cell signaling are also mediated by trans-membrane proteins through ligand binding [11].

Integrins are a class of highly important adhesion molecules that traverse the membranes of cells and are known to mediate cellular activities like cell-cell signaling, transport, cell adhesion, apoptosis, angiogenesis, and several other cellular activities [9, 12]. They are made of two sub-units designated α and β. Active interactions are fostered by the formation of noncovalent complexes between integrins and certain molecules found in the extracellular matrix (ECM) or on the surface of nearby cells. The putative recognition motif through which integrins bind some of their targets is the RGD amino acid sequence on their ligands.

Integrins represent one of the four major classes of protein receptors that are involved in cell-cell and cell-matrix interactions. The other three are: cadherins, selectins and members of the immunoglobulin superfamily [13]. Many of the cellular processes mediated by integrins are fostered by noncovalent interactions between integrins and extracellular matrix proteins (ECM) containing the RGD (Arg-Gly-Asp) recognition motif [14, 15]. This characteristic has made the

integrin-RGD recognition system a very important one for studying integrin involvement in cellular processes such as adhesion, signaling, differentiation, and apoptosis. In addition, because of the roles of integrins in diseases such as cancer, osteoporosis, and thrombosis, the integrin-RGD recognition system has become a major target of interest for developing peptidomimetic drugs and pharmaceuticals. For example, some classes of integrins have been found to be over-expressed in tumor cells, thereby making them attractive targets for anti-tumor agents [16, 17].

Many studies have been carried out to investigate the modality by which RGD-based peptides interact with integrins; with the main goal of inhibiting or preventing unfavorable integrin interactions believed to be detrimental to normal cell life. For example, both $\alpha_v\beta_3$ and $\alpha_v\beta_5$ integrins have been found to play major roles in angiogenesis, a process through which tumor growth is effected via the formation of new blood vessels [18, 19]. Peptide inhibition studies have shown that monoclonal antibodies and other small molecules interact with these receptors and consequently inhibit the angiogenic process. A good example is the methylated cyclic RGD peptide, Cilengitide (cyclo(-RGDf[NMe]V-)), which has now entered phase II clinical trials [20, 21].

There is significant variability in the manner in which different integrins interact with RGD-containing proteins or ligands. For example, by varying the identity of amino acids flanking the RGD sequence in a ligand, selectivity can be achieved between $\alpha_v\beta_3$ and $\alpha_{IIb}\beta_3$, both of which belong to the same β_3 integrin subfamily [22, 23]. Studies of the crystal structures of the extracellular domains of integrins, as well as mutagenesis of binding regions, have greatly improved the understanding of how integrins bind to their ligands. However, due to the fact that three dimensional structures of most of the integrin subtypes are still not available, it remains a challenging problem to determine what factors control specificity between different integrin/ligand pairs [24]. More techniques are thus needed to gain a better understanding in this area. It is interesting that many of the studies in the area of integrin recognition sequence have hinged on the successful reduction of the huge ligands to small recognition sequences [9]. In other words, it may be possible to delineate the interactions between integrins and their macromolecular ligands

by using short peptide sequences that represent the primary interaction sites between the receptor and the ligand.

2.2.1 The Integrin Structural Make-up

Integrins are made up of two subunits (α and β) noncovalently held together. Each subunit comprises a cytoplasmic portion, a transmembrane portion, and an extracellular portion. The extracellular domain is about 700-1100 residues long and the cytoplasmic region is about 30-50 residues long. Findings from electron microscopy and similar studies reveal that the entire integrin molecule is about 28 nm in length, having an oblong (or globular for some types) head of about 8-12 nm in diameter and flexible tails of about 18-20 nm in length [25]. It is believed that the head region is important for the association of the subunits (i.e. heterodimer formation). The individual morphology of each subunit has also been studied to an appreciable extent. Primary sequence analyses have shown that the N-terminal domain of all α subunits comprises about seven repeats that are very similar in sequence and each containing about 60 amino acids. An A-domain-like segment (similar to the A-domain of the von Willebrand factor) has also been identified in both the α and the β subunits of certain integrins (referred to as αA- and βA-domains respectively).

The only region of an integrin molecule for which 3-dimensional structure has been obtained so far is the αA-domain (also known as the I-domain). This a region of about 200 amino acids that is found in many proteins involved in adhesion mediation, including integrins. The crystal structure of the A-domain of an α subunit of integrin $\alpha_M\beta_2$ (considered to be the major integrin of phagocytic cells), a member of the β_2 integrin subfamily (whose main functions are the mediation of the adhesion of leukocytes to endothelium and the subsequent transport across the endothelium into inflamed organs) was obtained by crystallization in the presence of Mg^{2+} and solved using multi-wavelength anomalous dispersion (MAD) [26]. The X-ray structure of the same

domain was also independently obtained in the absence of any metal ion by Baldwin *et al*. [27]. Both results indicate the presence of a cation-binding motif in the αA-domain that is referred to as a metal ion dependent adhesion site (MIDAS). In the case of the A-domain of α_M mentioned above, the MIDAS motif is found in the region containing residues 140-144 and comprises a well conserved DXSXS sequence (where X is any amino acid but is G in the case of α_M A-domain). This short sequence provides three coordination sites for the metal ion (typically Mg^{2+}, Ca^{2+}, or Mn^{2+}). A threonine at position 209 and an aspartate in position 242 are also part of the motif, each contributing an additional coordination site for the metal ion. A sixth coordination site is believed to be provided by the side chain of a glutamate residue on an adjacent molecule. Among the three dications mentioned above, Mn^{2+} is found to bind the strongest while Ca^{2+} binds the weakest. Mg^{2+} has about 10 fold lower affinity compared to Mn^{2+} [27]. A similar cation binding motif is also found on all integrin β subunits [28-32]. It is proposed that this MIDAS region on the β subunits is also located in the A-domain-like portion of β subunits (the βA-domain). It is believed that the MIDAS region can play a role in the integrin ligand binding process either by direct coordination of an acidic residue (e.g. glutamate or aspartate) or indirectly via allosteric effects resulting from the cation binding.

Owing to the lack of crystal structures for the whole integrin molecule, much of the remaining portion of the molecules, (both α and β subunits, except the αA-domain) have only been studied using computational structure prediction models. The N-terminal portion of α subunits has been predicted to have a β-propeller fold [33]. This domain is made up of seven sequence repeats of about 60 amino acids each. In this β-propeller model, the A-domain of the α subunit is believed to be tethered to the upper face of the β-propeller domain by a hinge and the Ca^{2+} binding EF-hand-like motifs are believed to be on the lower face. The C-terminal end of α subunits is predicted to have a β-strand configuration with several β-sandwich and β-barrel modules [34].

The N-terminal portion of β subunits is predicted to adopt a α-helix-based morphology. It is composed of a 90-amino-acid repeat having about seven cysteine residues. There is remarkable similarity among the integrins in terms of the amino acid sequences of their β subunits. The β subunits also share similar structure, each typically comprising a large extracellular domain containing ligand recognition sequence(s) and four cysteine-rich repeats, a transmembrane domain, and a short cytoplasmic tail at the C-terminus [35]. The occurrence of these cysteine-rich repeats in the extracellular domain of the integrin β subunits led researchers to link the structure of this domain to that of the epidermal-growth-factor (EGF)-like modules in laminin, which only differs in the number of cysteine residues in its first repeat. As mentioned earlier, the A-domain of the β subunits of all integrins also contains a metal ion binding MIDAS region.

It should be noted that not all the regions of the integrin molecule that exhibit peculiarities in their folding pattern or other features participate in ligand binding. Studies using truncated integrin subunits have revealed that the binding site in integrin molecules lacking an αA-domain is located in the head region, comprising the N-terminal portions of both integrin subunits. For integrin molecules that have αA-domains, the major binding site is situated within the I-domain [25]. Supporting evidences for this conclusion include the mapping of anti-functional monoclonal antibodies to this domain [36-38], the ability of this domain to independently bind ligand when isolated, having binding constants (as determined by surface plasmon resonance) nearing those reported for intact integrin [39], as well as the recently solved crystal structure of a complex between the A-domain of α2 and a collagen peptide bearing a GFOGER motif [40].

As stated earlier, the MIDAS region is expected (in principle) to play a role in the integrin ligand binding process either by direct coordination between the metal ion (Mg^{2+}) and an acidic residue (e.g. glutamate or aspartate) or by an indirect allosteric mechanism upon binding of Mg^{2+} to this region. Evidence supporting this postulate is available from site-directed mutagenesis

experiments [41]. It has been reported in connection with the independent binding capability of αA-domains of integrins that there is considerable cation dependence in the binding process. In one study, Mn^{2+} was found to bind to the αA-domain of αM with high affinity [42]. It was also observed that other dications such as Mg^{2+}, Ni^{2+}, Zn^{2+}, and Co^{2+} bind this domain strongly as well, while Ca^{2+} and Ba^{2+} did not bind well. It is believed that cation binding helps to stabilize the domain during the binding process. The poor binding of Ca^{2+} and Ba^{2+} may be attributed to their sizes. Compared to their smaller counterparts, larger ions require more coordination sites, which may not be available from the αA-domain. It is possible that these larger ions will actually play an inhibitory role on ligand binding if they are present close to the binding site.

In order to gain better insight as to the relative contribution of the different regions of the integrin molecule, and the ligand itself, to the binding process, it is important to define the atomic contacts made between integrins and ligands. Such detailed elucidation of integrin-ligand contact points is only possible if the tertiary structure of an intact integrin molecule complexed with a ligand is available. The lack of such structures have necessitated the use of different techniques such as epitope mapping, mutation, ligand cross linking, and generation of a peptide fragment that comprises the binding site. There are clear advantages and disadvantages peculiar to each of these methods. For example mutation of residues (although capable of revealing key features of the binding pocket when the participating residues are removed) may cause other unwanted disturbances in the structure of the molecule. Also, generating chimaeras (ligand-binding peptide fragments) does help streamline the binding region on the integrin; however, there is possibility of missing important features of the binding pocket when using this approach. In light of the afore-mentioned, it is apposite to state that complementary (if not more robust and more efficient) techniques are needed in order to further probe the mechanisms of interactions between integrin proteins and ligands.

2.2.2 Integrin Ligand Binding

One of the most striking features of many integrin molecules is their ability to recognize multiple ligands. A list of known ligands recognized by integrin molecules is shown in Table 2.2 [9]. The crucial role of integrins as mediators of cell-cell interactions is highlighted in the preponderance of ligand types that are themselves receptors found on cell surfaces. Several microorganisms also bind to integrins and utilize this mechanism to gain entry into the cell.

Table 2.2 Common integrin extracellular ligands

Ligand	Integrin
Adenovirus penton base protein	$\alpha_v\beta_3$, $\alpha_v\beta_5$
Bone sialoprotein	$\alpha_v\beta_3$, $\alpha_v\beta_5$
Borrelia burgdorferi	$\alpha_{IIb}\beta_3$
Candida albicans	$\alpha_M\beta_2$
Collagens	$\alpha_1\beta_1$, $\alpha_1\beta_1$, $\alpha_{11}\beta_1$, $\alpha_{Ib}\beta_3$
Denatured collagen	$\alpha_5\beta_1$, $\alpha_v\beta_3$, $\alpha_{IIb}\beta_3$
Cytotact6in/tenascin- C	$\alpha_8\beta_1$ $\alpha_9\beta_1$, $\alpha_v\beta_3$, $\alpha_v\beta_6$
Decorsin	$\alpha_{IIb}\beta_3$
Disintegrins	$\alpha_v\beta_3$, $\alpha_{IIb}\beta_3$
E cadherin	$\alpha_E\beta_7$
Echovirus 1	$\alpha_2\beta_1$
Epiligrin	$\alpha_3\beta_1$
Factor X	$\alpha_M\beta_2$
Fibronectin	$\alpha_2\beta_1$, $\alpha_3\beta_1$, $\alpha_4\beta_1$, $\alpha_4\beta_7$, $\alpha_5\beta_1$, $\alpha_8\beta_1$, $\alpha_v\beta_1$, $\alpha_v\beta_3$, $\alpha_v\beta_5$, $\alpha_v\beta_6$, $\alpha_v\beta_8$, $\alpha_{IIb}B_3$
Fibrinogen	$\alpha_5\beta_1$, $\alpha_M\beta_2$, $\alpha_v\beta_3$, $\alpha_x\beta_2$, $\alpha_{IIb}B_3$
HIV Tat protein	$\alpha_v\beta_3$, $\alpha_v\beta_5$
iC3b	$\alpha_M\beta_2$, $\alpha_x\beta_2$
ICAM-1	$\alpha_L\beta_2$, $\alpha_M\beta_2$
ICAM-2,3,4,5	$\alpha_L\beta_2$
Invasin	$\alpha_3\beta_1$, $\alpha_4\beta_1$, $\alpha_5\beta_1$, $\alpha_6\beta_1$
Laminin	$\alpha_1\beta_1$, $\alpha_2\beta_1$, $\alpha_6\beta_1$, $\alpha_7\beta_1$, $\alpha_6\beta_4$, $\alpha_v\beta_3$
MAdCAM-1	$\alpha_4\beta_7$
Matrix metalloproteinase-2	$\alpha_v\beta_3$
Neutrophil inhibitory factor	$\alpha_M\beta_2$
Osteopontin	$\alpha_v\beta_3$
Plasminogen	$\alpha_{IIb}\beta_3$
Prothrombin	$\alpha_v\beta_3$, $\alpha_{IIb}\beta_3$
Sperm fertilin	$\alpha_6\beta_1$
Thrombospondin	$\alpha_3\beta_1$, $\alpha_v\beta_3$, $\alpha_{IIb}\beta_3$
VCAM-1	$\alpha_4\beta_1$, $\alpha_4\beta_7$
Vitronectin	$\alpha_v\beta_1$, $\alpha_v\beta_3$, $\alpha_v\beta_5$, $\alpha_{IIb}\beta_3$
von Willebrand factor	$\alpha_v\beta_3$, $\alpha_{IIb}\beta_3$

(Adapted from reference [9])

The putative recognition sequence (on the ligand) for many integrins is the short RGD amino acid sequence. Particularly, both $\alpha_{IIb}\beta_3$ and $\alpha_v\beta_3$ have been shown to bind their ligands via this motif. In fact, all the four extra-cellular matrix proteins that are known to bind $\alpha_{IIb}\beta_3$ (fibronectin, vitronectin, fibrinogen, and von Willebrand factor) possess this recognition sequence. It is a widely accepted notion that integrins bind such ligands with this RGD sequence via a direct interaction of an integrin-bound cation (in the MIDAS region) with the aspartic acid residue on the ligand [43]. Although many integrins recognize their ligands through this RGD motif, several others utilize different sequences. Moreover, RGD-specific integrin receptors are able to selectively bind one ligand in preference to another. The factors that control the selectivity of an integrin molecule towards its ligands include the affinity of the ligand, its conformational state, and characteristic features of the binding site on the integrin. Table 2.3 [9] has a list of some known integrin recognition sequences.

Table 2.3 Integrin recognition sequences

Recognition sequence	Ligand	Integrin
RGD	Adenovirus penton base protein, bone sialoprotein, collagen, decorsin, disintegrins, fibrinogen, fibronectin, prothrombin, tenascin, thrombospondin, vitronectin, von Willebrand factor	$\alpha_3\beta_1$, $\alpha_5\beta_1$, $\alpha_8\beta_1$, $\alpha_v\beta_1$, $\alpha_v\beta_3$, $\alpha_v\beta_5$, $\alpha_v\beta_6$, $\alpha_{IIb}\beta_3$
HHLGGAKQAGDV	γ - Chain of fibrinogen	$\alpha_{IIb}\beta_3$
GPR	α - Chain of fibrinogen	$\alpha_x\beta_2$
P1 peptide	γ - Chain of fibrinogen	$\alpha_M\beta_2$
P2 peptide	γ - Chain of fibrinogen	$\alpha_M\beta_2$
AEIDGIEL	Tenascin	$\alpha_9\beta_1$
QIDS	VCAM-1	$\alpha_4\beta_1$
LDT	MAdCAM-1	$\alpha_4\beta_7$
CS-1 peptide	fibronectin	$\alpha_4\beta_1$, $\alpha_4\beta_7$
CS-5 peptide	fibronectin	$\alpha_4\beta_1$
IDAPS	fibronectin	$\alpha_4\beta_1$
ICAM peptides	ICAM-1, -2, -3	$\alpha_L\beta_2$, $\alpha_M\beta_2$
DLXXL	Tenascin	$\alpha_v\beta_6$
GFOGER[α]	Collagen	$\alpha_1\beta_1$, $\alpha_2\beta_1$

[α] O, hydroxyproline

The presence of divalent cations has been shown to influence integrin ligand binding in three different ways. Some cations, when present in certain sites on the integrin molecule, are able to promote ligand binding while others may inhibit binding or induce selectivity of the receptor towards ligands [44, 45]. Surface plasmon resonance has revealed that certain integrin subunits (β_3) contain two distinct cation binding sites dubbed "ligand competent" (LC) sites and "I" sites. The LC sites must be occupied in order for the ligand to bind (i.e. these sites promote ligand binding); while cation binding to the I-sites inhibits ligand binding. The MIDAS region previously discussed will be an example of an LC site. The inhibitory I-sites have been shown to exhibit specificity for Ca^{2+}, similar to the EF-hand like domain earlier mentioned. It is concluded that the coordination between the LC sites and the I-sites control the ligand binding process [9]. Although the effects of cations in ligand binding by integrins have been well studied, it is still unclear whether or not these effects have any physiological relevance. It has been established, however, that the Ca^{2+} binding site (I-site) on $\alpha_v\beta_3$ integrin may modulate the adhesion of osteoclasts to bone surfaces, a process necessary for bone resorption.

It was previously mentioned in section 2.2.1 that the I-domains of integrins whose α subunits have such domains (about 9 of them) typically contain the binding region on those subunits; however, other regions of the α subunits (such as the EF-hand regions in $\alpha_L\beta_2$ and $\alpha_2\beta_1$) have also been shown to participate in ligand recognition [46, 47]. For these integrins, the direct involvement of the β subunit is not yet well established even though site-directed mutagenesis has indicated possible roles [48]. In the case of integrins whose α subunits lack I-domains, the ligand contact sites have been narrowed down to the N-terminal regions of both subunits. For example, in $\alpha_{IIb}\beta_3$, the seven N-terminal repeats of the α_{IIb} and the region containing residues 95-400 in β_3 have been identified as ligand recognition sites [49-51]. Based on the β-propeller model for the N-terminal portion of α_{IIb}, and in congruence with the notion that proteins

with β-propeller folds typically have their active sites at the top of the β-propeller, it is suggested that the α_{IIb} residues Gly-184 to Gly-193 and α_{IIb} Asp 224 (which the β-propeller model of α_{IIb} predicts to be at the top of the propeller) are involved in ligand binding by this receptor. Although site directed mutagenesis corroborates this hypothesis [52, 53], it is worthy of note that there appears to be some contradiction with data that suggest the contribution of the cation-binding EF-hand like motifs in α_{IIb} [54] to ligand binding as the β-propeller model places this region on the lower face of the propeller (*vide supra*), away from the binding region on the upper surface of the propeller [9]. Nevertheless, further support for the hypothesis that the β-propeller of integrins lacking I-domains bears the ligand recognition sites on its upper surface is found in the $\alpha_v\beta_3$, where the crystal structure of the N-terminal region of the whole integrin molecule complexed with an RGD ligand shows a direct interaction of the arginine residue on the ligand with the β-propeller [55]. In this solved crystal structure, it was observed that Arg and Asp side chains point in opposite direction and are in contact with the β-propeller and the βA-domains, respectively. The Arg of the ligand interacts with two aspartate residues (Asp 218 and Asp 150) in a small groove at the upper surface of the β-propeller of the α_V; while the ligand Asp interacts via its carboxylate group with a Mn^{2+} ion at the MIDAS region in βA-domain of the β_3. The Gly residue in the RGD occupies the interface between the two subunits and is involved in several hydrophobic interactions with the α_V subunit. These observations highlight a somewhat better defined role of the βsubunit in ligand binding for integrins lacking I-domains in comparison to their I-domain-containing counterparts.

Furthermore, certain synthetic peptides mimicking putative ligand-binding sites in integrins have been tested for their ability to bind ligand directly, and to inhibit ligand binding to integrin. For example, in an effort to determine the potential ligand binding site for fibrinogen on $\alpha_{IIb}\beta_3$, Charo et al tested several synthetic peptides (taken from the NH_2-terminal of βA-domain of the β3) for their abilities to inhibit the binding of fibrinogen to purified $\alpha_{IIb}\beta_3$ [56]. Two of the tested

peptides (residues 204-222 and 211-229) were observed to exhibit this inhibitory effect on the binding of fibrinogen to purified $\alpha_{IIb}\beta_3$ immobilized onto the bottom of Microtitre plates. The overlapping region between these two peptides has the amino acid sequence SVSRNRDAPEGG-NH_2 (residues 211-222), and blocked the binding of fibronectin, von Willebrand factor, and vitronectin to purified $\alpha_{IIb}\beta_3$. The potency of this overlap region was significantly greater than that of either of the two peptides comprised of amino acid sequences 204-222 and 211-229. The entire 204-229 segment on the β_3 subunit was also demonstrated to bind directly with fibronectin in an enzyme-linked immunosorbent assay. The results clearly indicate the crucial role of the overlap region 211-222 of the β_3 A-domain, and further buttress the participation of the β_3 subunit in ligand binding especially with integrins without αA-domains.

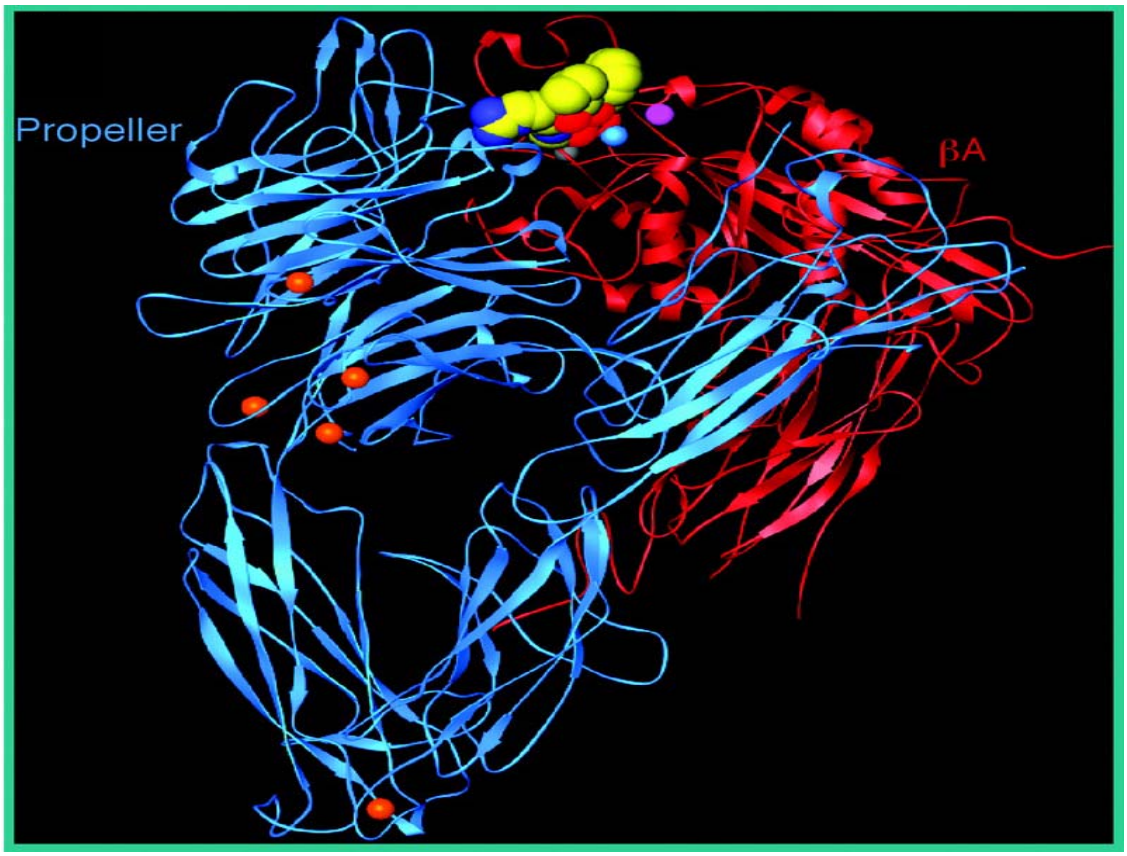

Figure 2.1 $\alpha_V\beta_3$ heterodimer bound to an RGD-containing cyclopeptide (cyclo (-RGDf[NMe]V-)) [55]

2.2.3 Analysis of host-guest complexation between integrins and RGD Ligands

Traditionally, host-guest complexation studies in biological systems are carried out using such techniques as calorimetry, centrifugation, UV-VIS, fluorescence, NMR, and X-Ray crystallography [57]. With respect to integrins, probing host-guest interactions with ligands have typically been carried out via immunoassay techniques. For example, information on the binding affinity of fibrinogen to $\alpha_{IIb}\beta_3$ immobilized on plastic support has been obtained from enzyme-linked immunosorbent assays (with K_d of 12nM) [56]. Similar techniques have also been used to evaluate the potency of certain ligands in inhibiting the binding of other competing ligands (measured as IC_{50} values) to certain hosts [58]. Site-directed mutagenesis and X-Ray crystallography have also been employed for pin-pointing the ligand contact points in the binding pocket of the receptor. Moreover, the bulk of information available to us today pertaining to the elucidation of binding properties of integrins to their native ligands as well as synthetic ligands has been obtained from a combination of these chemical and biochemical techniques.

Each of the methods mentioned above has its advantages and disadvantages. Considering their limitations, immunoassay development can sometimes be laborious and involves multiple steps. Centrifugation, UV-VIS, and fluorescence provide little information about the binding stoichiometry. The limitation of NMR is low sensitivity; it therefore requires greater sample quantity. This may become an issue when analyzing biological molecules since they are often available in only small quantities. The limitation of X-Ray crystallography is that the analyte complex has to be crystallized. In addition, both NMR and X-Ray crystallography are time consuming and have limited throughput. These shortfalls necessitate the development of new techniques for carrying out these types of studies with small sample quantities in a more efficient manner. Mass spectrometry based techniques have been demonstrated to aptly fit this task [59-63].

Since the report of D'Souza et al on the use of ESI-MS to detect the binding of a peptide segment of β3 (residues 118-131) to an RGD-containing ligand (RGDF) [44], there has been very

little work done in the application of mass spectrometry for studying molecular recognition in the integrin-RGD binding system. One of the goals of this work is to explore the use of molecular recognition mass spectrometry, employing an electrospray ionization source hyphenated to a quadrupole ion trap, as a complementary tool to investigate integrin-ligand interactions and to extract relative quantitative binding information for the noncovalent complexes between selected integrin fragments and RGD-containing peptides. The advantages of this kind of approach include greater speed, high-throughput screening capability and the direct determination of binding stoichiometry. The shortcoming in this method is mainly the limitation in mass analysis, given the high molecular weight of integrin molecules. Nevertheless, meaningful results can still be obtained from this method by studying binding interactions between peptide segments that are representative of the putative binding sites on the intact integrin molecules and short RGD peptide sequences from the ligand. This approach is what is explored in this work. Of the 24 or so integrin heterodimers discovered so far, our attention in this work is focused on peptide fragments from the two most commonly studied integrin molecules, $\alpha_{IIb}\beta_3$ and $\alpha_V\beta_3$.

2.3 The Mechanism of Electrospray Ionization Mass Spectrometry

Electrospray ionization mass spectrometry is one of the most versatile analytical techniques available today for qualitative and quantitative analysis. Analyte types to which ESI-MS is amenable include but not limited to biological macromolecules, small pharmaceutical molecules, organic and inorganic compounds, as well as environmental and geological samples. Emerging areas also include forensic and medical applications.

2.3.1 The Electrospray Process

In ESI-MS, the solution of the analyte is sprayed through a short length of capillary into an enclosed ionization chamber by a strong electric field. The field is generated by applying a positive or negative electric potential, typically 3-5 kV across the tip of the capillary [7, 8, 64, 65]. As a result of this high potential difference, the ions in the emerging solution at the capillary tip experience a “pull” towards the counter electrode. This pulling force is counterbalanced by the surface tension of the liquid, leading to the projection of the meniscus of the emerging liquid and subsequently to the formation of the so called ‘Taylor cone’ at the capillary tip [65]. Charge density will be highest at the apex of this cone. When the charge density attains a high enough value to overcome the surface tension of the liquid, charged droplets will form from this apex in a dynamic fashion. Overall, as the spray exits the tip of the capillary, it is turned into a mist of small (μm diameter) highly charged droplets. The droplets are carried towards the vacuum region by a potential gradient resulting from the high voltage difference between the capillary tip and the counter electrode, and become smaller (nm diameter) as they traverse the atmospheric pressure spray chamber. At some point along the way, the droplet reaches a small enough size that the repulsive force as a result of the excess charges it carries is balanced by the surface tension of the liquid, i.e. the Raleigh stability limit is reached. The radius R of the droplet at this point, is related to the surface charge q_R, the permittivity due to vacuum ε_0, and the liquid surface tension γ by equation 1[66]:

$$q_R = 8\pi(\varepsilon_0 \gamma R^3)^{1/2} \quad (1)$$

Further solvent evaporation leads to droplets with radii below the Raleigh radius and subsequently droplet fission [66]. The resulting daughter droplets following fission may undergo similar size reduction and hence reach their Raleigh stability limit and undergo fission as well. Ultimately, ions are formed from these droplets and they are sampled into the mass analyzer to record a mass spectrum.

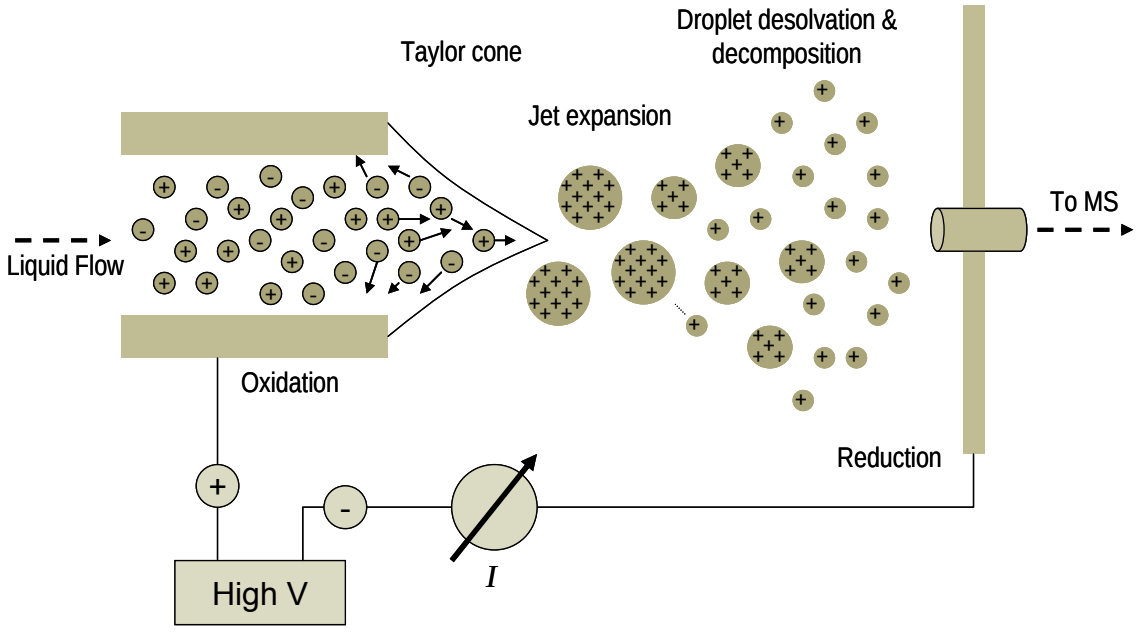

Figure 2.2 Schematic of the electrospray ionization process (positive mode). [67]

The mechanism of ion formation from the charged droplets has been a subject of debate over several decades. Two predominant mechanisms have been proposed. The charged residue model was proposed by Dole and co-workers and posits the formation of ions from very small droplets containing single ions that are themselves formed from the parent droplets that have undergone a series of fission events called 'Coulomb explosions'. Interspersed between these fission processes are moments of solvent evaporation. Gas-phase analyte ions are formed when the last solvent molecules evaporate from these single-ion-containing droplets and the single ions

now carry the excess charges originally borne by the droplets [68, 69]. Formation of such single-ion-containing droplets is deemed possible when the original concentration of the analyte is very low.

On the other hand, the ion evaporation model proposed by Iribarne and Thomson suggests a process of ion formation whereby solvated ions are lifted from the droplet surface as a result of the high electric field on the droplet surface as it shrinks due to solvent evaporation. This theory assumes the emission of the ions to be faster than the Raleigh disintegration [70, 71] and may occur at some droplet radius below 10 nm. This desorption of charge-carrying solvated ions from the charged droplet's surface in this manner will lead to a decrease in the excess charge on the original droplet and consequently suppress Raleigh explosions.

Both of these theories have been reviewed and subjected to several experimental tests. Although each theory has its own group of adherents who generally favor its postulates over those of the rival theory based on their own experimental observations, there seems to be a general consensus that formation of ions from charged droplets of large macromolecules is mainly governed by the charge residue model [66, 72-75] while smaller molecules may tend to form ions via the ion evaporation mechanism. By and large, the ultimate concern in the use of electrospray mass spectrometry is often not about the ion formation mechanism, as this seldom affects the resulting mass spectrum, but about the nature of, and modality by which different factors (both instrumental and otherwise) influence the observed mass spectrum. More importantly, the question of how well the observed ion intensities in the mass spectrum correlate with the solution phase equilibrium concentration/distribution of the species of interest still remains to be answered in a decisive manner. A lot of work is still on-going in this area and, an investigation into how instrument parameters and analyte parameters affect ionization efficiencies during electrospray mass spectrometry measurement constitutes the second major theme of this work and is discussed in more detail in chapters 5 and 6.

2.3.2 Important Consideration in the use of Electrospray Ionization Mass Spectrometry

It is pertinent to mention a few of the most important aspects that need to be considered in using ESI-MS for routine analyses. As mentioned earlier, a key question that is often asked is whether or not the ion abundances as observed in the mass spectrum are indeed representative of the solution phase concentrations of the analytes. For the most part, and in the absence of competing ions in solution, experimental evidences point to the fact that there is reasonable correlation between electrospray mass spectral ion abundances and the solution phase concentration of analytes. This is especially true for stable ions such as the singly charged alkali ions (Li^+ and the rest of the group) which usually already carry permanent charges in solution [66]. Aberrations from this norm usually arise when the ions being studied are in form of protonated or deprotonated species [66]. Here, the effects of order of basicity (or acidity) in the solution phase compared to the gas phase may be significant; obviously due to the marked difference between the solution phase and the solvent-free gas-phase environments. Deviation may also occur when solutions containing multiple species are sprayed. Multiplicity of species in the sprayed solution often leads to competitive ionization during the electrospray process (in the case of species that do not already carry a charge in solution) or ion suppression (in the case where certain ions have greater surface activities than ions of interest). Competition for ionization, as well as ion suppression, may be due to impurities in the solvent or electrolytes and buffer solutions introduced deliberately (or otherwise) from other processes such as chromatographic separations. Lastly, and of major interest to many researchers, is when ESI-MS is used to determine noncovalent binding affinities between a host and a guest. For the simple fact that it is not possible to determine *a priori* the solution phase equilibrium concentration of the host-guest complex for which binding affinity is not previously known, it is difficult to unequivocally correlate the gas phase ion abundance of the complex directly with its solution phase concentration.

One of the most important points to consider while using ESI-MS is the choice of solvent. Solvent properties such as polarity, viscosity, and dielectric constant, may evoke marked effects

on the electrospray process and alter the charging as well as charge state distribution of analytes. Polar solvents (generally with high dielectric constant) have the capacity to enhance charge separation. As such, ionization by analyte dissociation in solution will be promoted. Wang et al reported a consistent shift of charge state distribution towards higher states in their analyses of diquaternary ammonium salts when more polar solvents were employed [76]. These salts dissociate in solution to yield either singly charged (losing one of the two anions) or doubly charged (after losing both anions). Diquaternary ammonium chloride was analyzed in alcohol solvents methanol, ethanol, 1-propanol, and 1-butanol, (given in decreasing order of polarity). There was a marked decrease in the intensity for the singly charged species $[M\text{-}Cl]^+$ *pari passu* with an increase in intensity of the doubly charged species $[M\text{-}2Cl]^{2+}$ [76]. Similar effects on charge state distribution was observed when the same analyses were repeated with chlorinated solvents of decreasing polarity (CH_2Cl_2, $CHCl_3$, and CCl_4). The ratio of doubly charged to singly charged species increased going from CH_2Cl_2 to CCl_4. Another striking observation is the general decrease in total ion abundances with decreasing solvent polarity.

It is conceivable that other solvent properties such as surface tension, volatility, and viscosity may also influence the analyte charging process and may determine to a large extent how well the gas phase ion abundances mirror the solution phase distribution/concentration. However, effects of solvent polarity are usually more dominant and exhibit greater influence on analyte charging as well as charge state distribution compared to these other properties [77]. Nevertheless, a brief mention is made here in an attempt to portray a more complete picture of the influence of solvents properties in general. For example, solvent volatility will obviously influence the rate of solvent evaporation from the droplets during the electrospray process, a process critical to ionization efficiency [73]. Highly volatile solvents will exhibit higher rates of evaporation and thus shrink the droplets faster during the electrospray process. This phenomenon results in higher ion abundance in the case of small ions where ion evaporation is considered the dominant ion formation mechanism [73]. Charge state distribution may also be

altered since high rate of evaporation will shorten the period of time for which a low surface charge density is maintained on the droplet surface, leading to a decrease in the proportion of low charge state ions that are able to get desorbed from the droplet surface (via ion evaporation) as solvent evaporates. Also, droplets of smaller size are produced when spraying solvents that have less surface tension, thus leading to better sensitivity. Effects of solvent types on analyte cluster formation, structure and stability have also been reported [78]. Solvents of low dielectric constant and high viscosity tend to favor cluster formation in solution phase, but, as noted above, may have negative effects on ionization efficiency.

Among other parameters to be considered in an ESI-MS experiment are ionization mode, solvent flow rate, and capillary size. It is a known fact that positive ionization mode is usually more sensitive than negative mode. The lower sensitivity of negative ionization mode have been attributed to less stability of the Taylor cone as well as high propensity of corona discharge during the electrospray process. Wu et al [79] have reported a method of improving the overall signal quality of negative mode ESI-MS measurements. In their report, a group of additives were evaluated for their effects on the negative mode signals of selected analytes. Formaldehyde was found to improve analyte signal tremendously compared to when no additive was added.

The flow rates employed in typical ESI-MS experiments are usually in the low microliter per minute range (1-10 μl/min). Higher flow rates (ranging from a few tens of microliters per minute to low milliliter per minute) may also be encountered when ESI is coupled to liquid chromatography. By inspecting spectra of different analytes in our laboratory, it is observed that there is usually a marked variation between spectra acquired at low flow rates and those acquired at high flow rates for the same analyte solution. One of the most common differences is in charge state distribution of analyte species in the spectra observed. Although higher signal intensities can sometimes be obtained with high flow rates when a stream of drying gas is flown co-axially with the liquid jet exiting the spray capillary, there is greater tendency for ion suppression from electrolyte ions which are usually present in greater concentration than the analyte of interest

when high flow rates are used in ESI-MS experiments. Better signals and higher sensitivity are usually obtained at low flow rates due to improved ionization efficiency and reduced ion suppression. This is the basis of the latest version of ESI commonly referred to as 'nano-ESI', where flow rates as low as a few tens of nanoliters per minute are employed. Covey [80] has reasoned that, under this mode of operation of ESI-MS, the response is proportional to the amount of material present. Aside the obvious advantage of lower sample consumption when ESI-MS is operated in the low flow rate regime, there is overall improvement in ionization and ion transmission efficiencies. This can be attributed to the smaller droplets formed when low flow rates are used. The smaller droplets would have higher surface/volume ratios and hence enhance the desorption efficiencies of ions [81, 82]. Solvent evaporation is also more efficient with smaller droplets; leading to enhanced ion emission or droplet fission (since the Raleigh limit is reached more quickly). This whole process is further enhanced when narrower bore capillaries are employed. Capillaries with internal diameters in the range 1-2 μm are typically employed for low flow rates applications. These factors have accounted to the improved limit of detection in nano-ESI compared to conventional ESI-MS [83, 84].

Another aspect of ESI-MS worthy of mentioning, particularly with regards to its application in noncovalent studies, is the low internal energy imparted to the ions during the electrospray process compared to other mass spectrometry techniques. This accounts for the very low degree of fragmentation encountered in ESI-MS. This is particularly of greater importance in studies of noncovalent interactions considering the need to preserve the noncovalently bound host-guest complex during its transfer from solution to gas phase. The energy imparted to the ion during the electrospray process can be largely controlled by varying the voltage difference between the cone and the skimmer. Higher voltage difference between these two points has been shown to impart higher internal energies, and consequently shifting the charge state distribution towards lower values [76, 85, 86].

2.4 Application of Electrospray Ionization Mass Spectrometry to Molecular Recognition

Since the introduction of the Electrospray Ionization Mass Spectrometry (ESI-MS) in the late 1980s by John Fenn, the technique has become one of the most preferred methods for the analysis of biological macromolecules. The advantages of mass spectrometric techniques in general, and ESI-MS in particular, are many. The most notable are specificity, sensitivity, and speed; aptly described as the three "S" advantages by McLafferty [87]. More so, with regards to study of noncovalent interactions, ESI-MS is a "softer" ionization technique, and thus, can ensure the successful transfer of host-guest complexes from solution phase to gas phase without disrupting the noncovalent interactions between the host and the guest. It is also possible to determine host-guest binding stoichiometry directly from the analysis. Furthermore, in ESI-MS it is possible to evaluate both solution phase and gas phase binding affinities [88-104]. Solution phase techniques usually include titration experiments and competitive binding experiments. Examples of gas phase techniques are ion-molecule reactions, collision induced dissociation and thermal dissociation of isolated gas phase complex ions. Other more unique advantages of ESI-MS include its amenability to a wide-ranging spectrum of solvent types, and the occurrence of multiple charging that makes it possible to observe analyte species whose masses are normally outside the range of common mass analyzers. Characterized by these figures of merit, ESI-MS has gradually become the method of choice for probing noncovalent interactions in biomolecules involved in host-guest complex formation.

Although the advantages of ESI-MS for studying noncovalent interactions far outweigh the disadvantages, it is pertinent to mention a few of the draw backs. The most notable limitation of this technique is the (more often than not) lack of correlation between binding affinities measured in solution phase (with ESI-MS as the detection mechanism) versus those measured in gas phase, particularly for small molecule complexes. This lack of correlation is attributable to two main reasons. The first of them, the so called response factor problem, has to do with the inability to determine the response factor of the newly formed complex during the electrospray process,

which makes it difficult to directly correlate the equilibrium concentration of the complex to its ion abundance as observed in the mass spectrum. More details on response factor and the analyte parameters that influence it are presented in Chapter 4 (*vide infra*). The second major reason for discrepancies between solution phase and gas measurements is the inherent difference between the two environments. Charge-based noncovalent forces like electrostatic interactions are usually enhanced in absence of solvents, thereby contributing more to the binding than in solution phase.

2.4.1 Methods for Determining Binding Affinity via Electrospray Mass Spectrometry

Quite a good number of mass spectrometry-based methods have been established for determining binding affinities of host-guest complexes. Both solution phase measurements (where ESI-MS is used as a detector) and gas phase measurements (where the complex is isolated in the gas phase and dissociated by one of several available methods) have been reported [3, 6]. Solution phase methods include MS-based titration techniques [105-109], melting curves measurements [110] competitive binding analyses [111], and the newly introduced dynamic titration (a more efficient variant of MS titration) [112]. All of these methods rely on the ion abundances of host, guest, and complexes in the mass spectrum as a measure of the solution phase equilibrium concentrations of the species involved. Ion abundances of relevant species are then used in place of the equilibrium concentrations in the equilibrium expression to obtain either the association constant or the dissociation constant. It is important, while employing these methods, to ensure that what is being measured is actually solution phase and not gas phase parameters. Thus, the method should be set up in such a manner that no part of the complex dissociates during the transfer from solution phase to gas phase. Also, likely shift in the solution phase equilibria during the electrospray process should be taken into consideration. Solvent evaporation from the droplet surface (which may lead to pH variation) during the electrospray process is expected to affect the position of the equilibrium. However, a number of reports found in the literature point to the fact that minimal shift in the position of the equilibrium is

observed during the electrospray process. The most recent report in connection with this phenomenon was by Wortmann *et al* [113], who studied the dimerization equilibrium of fluorescent Rhodamine B-sulfonyl chloride during the electrospray process using laser-induced fluorescence (LIF) in conjunction with phase Doppler anemometry (PDA) [113]. Measurements were carried out at different spray positions and the droplet size was monitored via the PDA while fluorescence was monitored in the droplets via the LIF. From their results, Wortmann et al. concluded that there is no influence of decreasing droplet size on the monomer-dimer equilibrium. Although these authors mentioned some increase in concentration in the droplets when another analyte was studied in this manner, more evidence supporting the previous conclusion is found in the earlier work done by Wang and Agnes [114] on the complexation of strontium with EDTA (eq 2).

$$Sr^{2+} + EDTA^{4-} \underset{k_{-1}}{\overset{k}{\rightleftharpoons}} [Sr\text{-}EDTA]^{2-} \qquad \text{(eq 2)}$$

In this system, the forward reaction rate is in the order of $10^9 M^{-1} s^{-1}$ and the reverse rate is in the order of 1 s^{-1}. The ESI process is in the time scale of approximately 10^{-2} s [3]; much faster than the rate of dissociation of this complex. In spite of the expected shift (to the right) in the position the authors reported that the observed deviation in the equilibrium position was less than expected [3]. From the foregoing, it appears there is little or no variation in the equilibrium position going from solution phase to the gas phase during the electrospray process. It is possible that this effect is system-dependent; as such, care should still be taken to ensure that the system being investigated is not subject to pronounced shift in the solution equilibrium during the electrospray process. Nevertheless, measurement of solution phase equilibrium systems via ESI-MS is still justifiable though can be system-dependent in its success.

Perhaps the most common MS-based solution phase technique for evaluating dissociation constants of host-guest complexes is the titration technique. In this approach, the host concentration is held constant while that of the guest is varied over a range of concentrations falling within its linear dynamic range. The gas phase ion abundances of the noncovalent host-

guest complex and free (unbound) host can be fit to algorithms based from solution-phase host-guest association models to extract binding constants. Values are determined using graphical methods such as the well known Scatchard analysis [106] and similar semi-logarithmic plots [115, 116]. In the figure, [HG] is complex equilibrium concentration, $[H]_0$ is initial host concentration, [G] is equilibrium concentration of guest, K_D is the dissociation constant, n is the number of binding sites on host).

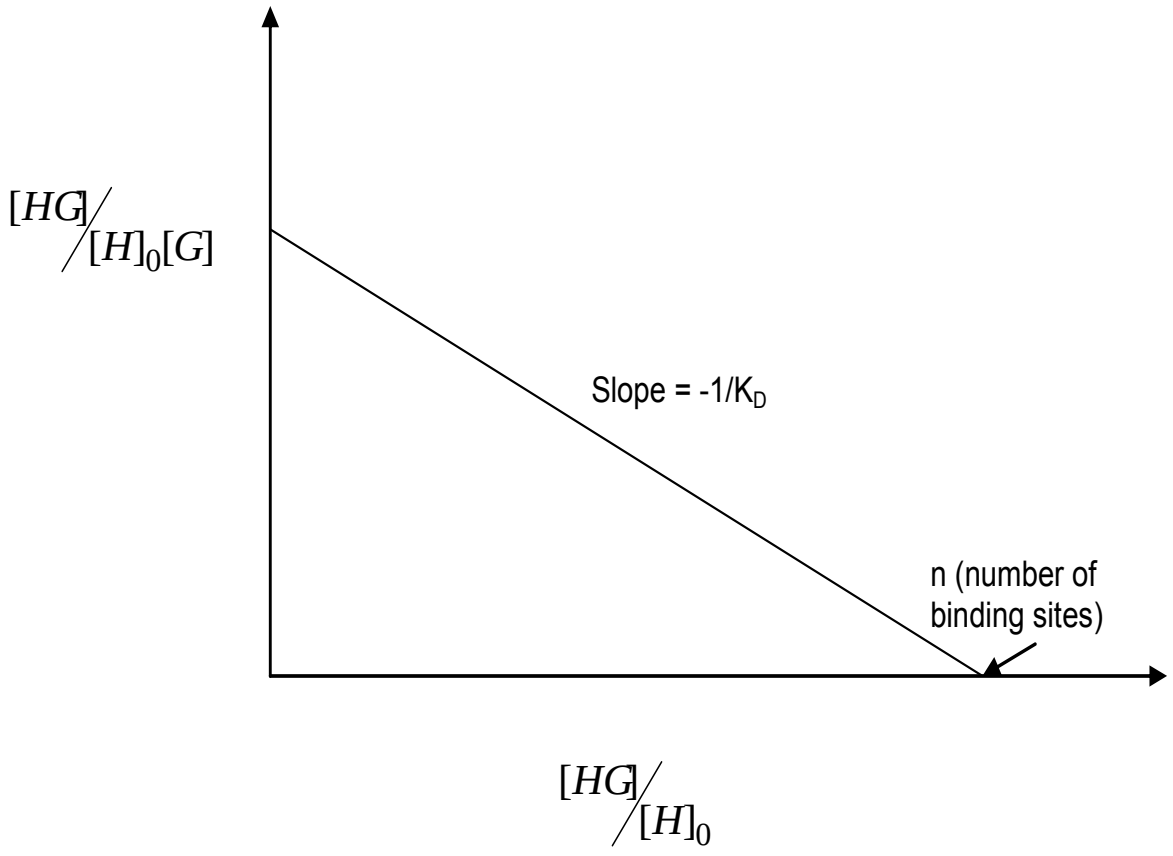

Figure 2.3 Schematic of a Scatchard Plot

This technique has been successfully employed to determine the binding affinity of vancomycin and ristocetin to the bacterial cell wall peptide “Ac-L-Lys (Ac)-D-Ala-D-Ala” by ESI-MS [106]. The calculated binding constant using this method was found to be in good agreement with those obtained from traditional solution phase methods. It is also possible to determine the binding stoichiometry using this method; a feat that is sometimes difficult to achieve using other traditional solution phase methods. This is one of the advantages of MS-based system over their traditional solution phase counterparts. Other successful applications of the titration technique for binding affinity determination have been reported by Greig et al., who used ion spray to study the interaction of bovine serum albumin (BSA) with oligonucleotides [107], and Loo et al. in their study of binding interactions of SH2 (a domain commonly found on oncoproteins and known to typically bind a phosphorylated tyrosine residue) motifs with phosphorylated and unphosphorylated peptides [90]. Semi-logarithmic methods have also been employed to extract binding affinity information from MS-based titration data. Using this method, Schug et al have demonstrated good agreement between relative binding constants derived by fitting mass spectrometric titration data to a quadratic algorithm and those obtained from isothermal calorimetry (ITC) measurements for a set of small molecule chiral recognition systems [116].

Another variation of the MS-based titration technique is the recently introduced dynamic titration technique by Frycak and Schug [112]. In this method, a small volume of guest solution with known concentration is injected into a flow of host solution and made to pass through a length of tubing where band-broadening takes place. This results in a continuous Gaussian concentration profile of the guest. The tubing is connected to the ESI-MS where the time-dependent intensities of the species are recorded. Using a modified Gaussian function that incorporates the amount of guest injected, the flow rate of the host solution, the width of the resulting Gaussian peak, and the time at which the guest concentration reaches its maximum during the band-broadening process (here taken as the time value that divides the area under the Gaussian peak into two equal halves assuming peak symmetry), the K_d is derived by fitting a

curve unto the Gaussian peak. The method was tested and demonstrated to work suitably well for noncovalent complexes of cinchona alkaloid chiral selectors with N-dinitrobenzoylleucine enantiomers and a series of cyclodextrins with sulfonated azo dyes. The advantages of this dynamic titration technique include speed and ability to carry out more replicate measurements efficiently with limited sample preparation. Also, less analyte is consumed in this method compared to conventional titration experiments, where discrete solutions of varying concentrations are titrated against a fixed concentration of host.

Another very commonly employed solution phase method for determining binding constant via ESI-MS is the competitive binding method. Two most popular variants of this method have been used extensively. In one approach, a number of different guest molecules compete for the binding site of the host (or *vice versa*). The relative peak intensity of the free host and the ensuing complexes are then measured on the mass spectrum and, based on the assumption that the ionization efficiencies of the host and the complexes are similar, the binding constant is determined from the peak intensities of the relevant species. This method was successfully employed by Jorgensen et al. and found to be suitable for calculating solution absolute as well as relative binding affinity of the antibiotics vancomycin and ristocetin with some short peptides [117]. A relation between the solution phase equilibrium concentrations of the free antibiotic and the complexes formed with three ligands is depicted in equation 3 below:

$$[A]_i = \frac{[A]_i * [A]_0}{(A + AL_1 + AL_2 + AL_3)}, \qquad \text{eq (3)}$$

where A_i is any form of the antibiotic, $[A]_0$ is the initial concentration of the antibiotic, L_1, L_2, and L_3 represent each of the three ligands, and AL_1, AL_2, and AL_3 are the peak intensities of the complexes. Using this equation, it is possible to relate the peak intensity of any species to its equilibrium concentration; binding constant can thus be derived directly. When an equimolar mixture of the antibiotic and the three ligands is used, the binding constant for any of the

complexes (shown for complex AL_1 in this case) in the system can be derived from the expression below:

$$K_{AL_1} = \frac{[AL_1]}{[A]*[L_1]}, \qquad \text{eq (4)}$$

Since $[L_1] = [L]_0 - [AL_1]$, and $[L]_0 = [A]_0$, then

$$[L_1] = [A]_0 - [AL_1] \quad , \qquad \text{eq (5)}$$

If $[A]_0 = [A] + [AL_1] + [AL_2] + [AL_3]$, then,

$$[L_1] = [A] + [AL_2] + [AL_3], \qquad \text{eq (6)}$$

Equation 4 then becomes

$$K_{AL_1} = \frac{[AL_1]}{[A]*([A]+[AL_2]+[AL_3])}, \qquad \text{eq (7)}$$

Similar expressions can be derived for any of the complexes. Results obtained from this method were found to be in good agreement with literature values. The draw back encountered with this method is the necessary assumption that ionization efficiencies of the free host and the complexes have to be similar. This is usually only true when studying systems of host-guest complexation where the host is considerably larger than the guest such that there is little or no difference between the physicochemical properties of the free form of the host and the guest-bound form.

The second variation of the competitive binding method was first employed for ESI-MS binding affinity determination by Kempen and Brodbelt [111]. It was an adaptation of the technique earlier used by Gokel et al. for determining binding constants between crown ethers and Ca^{2+} ions using ion selective electrodes [118]. A similar method was also applied in NMR studies of crown ether complexes with Rubidium ion [119]. The approach involves the use of a reference complex whose binding constant is known to determine the binding constant of a new complex whose host or guest component is also the host or guest component of the reference

complex. In other words, both reference and new complexes must have one species (host or guest) in common. This approach was successfully demonstrated by Kempen and Brodbelt for determining the log K values for complexation of several crown ether complexes with alkali metal cations as well as complexes involving dibenzo-16-crown-5 and its derivatives with sodium and potassium ions. They first obtained a calibration curve for the reference host-guest complex, and then introduced a host (or a guest as the case may be) into the solution containing the reference host-guest system, thereby establishing a competitive equilibrium. Since the binding constant of the reference complex is known, the concentration at any point in time can be calculated from the calibration curve. The binding constant of the new host-guest complex can thus be obtained from this information. The method was validated by comparing the results obtained with those reported in the literature from solution phase techniques applied for the same systems studied and the results were in good agreement.

The technique is straightforward and does not require any assumption about the response factor of any species since the intensity of the reference species is the only quantity monitored. It also does not require the new complex or the host or guest to be ionic, eliminating complications from inconsistent ionization efficiencies of these species. It does however require the ionization efficiency of the reference host to be fairly constant throughout the calibration range as well as when the new host or guest is introduced into the system. In other words, there should be little or no perturbation in the ionization efficiency of the reference complex due to the presence of the new host or guest. Another draw back is that it is only applicable to systems of known binding constants that have been studied with other solution phase techniques.

Binding affinities have also been studied by ESI-MS using gas phase techniques. Unlike their MS-based solution phase counterparts, MS-based gas phase methods do not generally present any concern with respect to ionization efficiencies of species of interest. The results are rarely in agreement with solution phase measurements, especially in absolute terms, except in few cases. The discrepancies are not unexpected given the fact that certain noncovalent forces

which are present in solution phase are completely absent in the gas phase (e.g. hydrophobic interactions), while many (such as electrostatic interactions) are actually enhanced in the solvent-free medium due to the fact that electrostatic forces are inversely proportional to the dielectric constant of the medium.

Some of the gas phase techniques for determining binding affinity with ESI-MS include cone voltage-induced dissociation [120], collision-induced dissociation (CID) [121, 122], blackbody infrared radiative dissociation (BIRD) [123, 124], heated capillary dissociation [125], and guided ion beam tandem mass spectrometry [126]. The most commonly used of these techniques are the cone voltage-induced dissociation and the collision induced dissociation. In the cone voltage-induced dissociation method (also called in-source CID), the kinetic energy of the complex species is increased by increasing the cone voltage. By increasing the cone voltage in a stepwise manner, the voltage VC_{50} required to dissociate 50% of the complex can be calculated. This technique has been employed by Rogniaux et al [127] in their study of enzyme-inhibitor interactions. Although it is possible to access the relative gas phase stabilities of various complexes with the VC_{50} method, there is no direct quantitative correlation between the VC_{50} values and the binding energy. Rogniaux et al. also noted the lack of correlation between the VC_{50} and the IC_{50} values (a measure of the inhibitory activity of a drug) of the studied inhibitors.

Another widely used method for evaluating gas phase binding affinity is the collision induced dissociation (CID). In the CID method, the complex ion is mass selected and then fragmented by collision with inert gaseous molecules in a collision cell. The resulting daughter ions are then scanned in a mass analyzer. The whole process can typically be carried out in a triple quadrupole mass analyzer, where the complex mass selection is carried out in the first quadrupole, the fragmentation in the second quadrupole, and the final mass analysis of fragment ions in the third quadrupole.

Cheng et al. [122] has described a collision model that can be used to calculate the relative internal energies transferred to the complex to induce dissociation during the CID

process. In their model, they have reasoned that the maximum collision energy that the complex ion can attain during the passage through the collision cell is the summation of all the center of mass energies of all the collisions taking place. This energy is influenced by a number of factors including the collision gas pressure, molecular weight of collision gas, and the cross section of the ions. The center of mass energy, E_{com}, for a single collision process can be defined as [84]:

$$E_{com} = E_0\left[\frac{m_2}{m_1 + m_2}\right], \qquad \text{(eq 8)}$$

where E_0 is laboratory frame kinetic energy of the ion injected into the collision cell, m_1 is the mass of the complex ion, m_2 is the mass of the target gas. When summed over the number of collisions taking place in the collision cell, the internal energy acquired by ions passing through the collision cell is described as [122]:

$$E_{\text{int}} = \phi\frac{m_2}{M}E_0\frac{m_1}{m_2 C_D}\left\{1 - \exp\left(-C_D n m_2 \sigma l \Big/ m_1\right)\right\}, \qquad \text{(eq 9)}$$

where E_{int} is the internal energy acquired by the ion, $M = m_1 + m_2$, ϕ is the average fraction of center of mass kinetic energy transferred to internal energy of the ion in a single collision, C_D is the drag coefficient, n is the gas number density, σ is the collision cross-section, and l is the length of the collision cell. When the exponential term in eq 9 is small and $\phi = 1$, eq 9 reduces to:

$$E_{\text{int}} = \left(\frac{m_t}{M}E_0\right) n\sigma l\,, \qquad \text{(eq 10)}$$

Eq 10 is simply the product of the center of mass energy per collision (equation 9) and the total number of collision.

The success of CID for binding affinity evaluation has been demonstrated by several groups using different mass analyzers for ion isolation, fragmentation, and analyses. Li et al. [128] have used tandem ESI-MS to study the binding affinities of rapamycin and four of its analogs to the cytoplasmic receptor FKBP. Their results showed correlation with solution phase measurements for the same complexes. ESI-FT-ICR has also been successfully used for CID

studies [129]. The voltage required to dissociate 50% of the complex ion, V_{50} (analogous to the VC_{50} for the cone voltage induced dissociation) can be obtained from this type of experiments by plotting the injection voltage vs the relative intensity of the daughter ion or, better still, the normalized percent parent complex ion remaining vs the collision voltage.

A special application of the CID method is the Cook's kinetic method for evaluating enantioselectivity excess of chiral compounds involved in host-guest complex formation. The underlying principle is the competitive dissociation of the trimeric complex of each of the enantiomers with a reference chiral compound and a metal ion when subjected to collision induced dissociation [130, 131]. The ratio of the two dissociation rates is a measure of the binding strength of the chiral compound to each enantiomer and can be used to access selectivity.

Other gas phase methods such as guided ion beam tandem mass spectrometry (GIBMS) [126], heated capillary dissociation [125], and blackbody infrared radiative dissociation (BIRD) [124] require special instrument design and, as such, are not as commonly used as CID methods.

CHAPTER 3

EXPERIMENTAL

3.1 Introduction

The experiments performed in this dissertation work can be broadly categorized into two classes based on the aims and objectives they serve to achieve. The first category of experiments addresses the unprecedented attempt to evaluate binding affinities between synthetic integrin peptide fragments and RGD peptides using data acquired from ESI-MS studies. The experiment in this category includes:

- Experiment 1 (Chapter 4): ESI-MS -based molecular recognition studies between Integrin Fragments and RGD-Based Peptides [132].

The second category of experiments pertains to the study of the effect of analyte physicochemical properties and instrumental parameters on the ESI-MS response of model analytes. The following experiments fall under this category:

- Experiment 2 (Chapter 5): Evaluation of the Effects of Physicochemical Parameters on the ESI-MS Response of GXG Tripeptides: An Approach Using Multivariate Statistical Analyses;
- Experiment 3 (Chapter 6): Chemometrics Study of the Influence of Instrumental Parameters on ESI-MS Analyte Response using Factorial Design.

3.2 Experiment 1: ESI-MS -based molecular recognition studies between Integrin Fragments and RGD-Based Peptides

The aim of this experiment was to study the binding characteristics of integrin peptide fragments with short RGD peptides using ESI-MS. Integrin peptide sequences were first identified which represent the binding pocket on the intact protein molecule. This information was extracted from literature reports and crystallographic data. The identified sequences were then synthesized and tested for their affinities towards short RGD-containing peptides.

All synthetic peptides used in this investigation (Tables 3.1 and 3.2) were supplied by a collaborator (Dr Jung-Mo Ahn, Department of Chemistry, UT-Dallas). The peptides, GRGDsP, RGD, and Glycoprotein IIb fragment 206-306 (TDVNGDGRHDL), were obtained commercially from Bachem California Inc (Torrance, CA). GRGDSP was obtained from Bachem Bioscience Inc (King of Prussia, PA) and GRGDNP was purchased from Biomol (Plymouth Meeting, PA). Stock solutions of the peptides were prepared by dissolving known amounts of the peptides in 50/50 methanol/water solutions. Different concentrations used for the titrations were prepared from the stock solutions by dilution with the same solvent mixture. Ammonium acetate buffer and acetic acid were likewise prepared from stock solutions and incorporated directly into the sample mixtures for analysis.

Table 3.1 Primary Sequences of studied host peptides and their masses

Peptide	Sequence	Mass (Da.)
Integrin α_{IIb} segment [α_{IIb}f; (184-193)]	Ac-GAPGGYYFLG-NH_2	1041.4
Integrin α_V segment [α_vf; (213-220)]	Ac-AQAIFDDSYLG-NH_2	1239.3
Integrin β_3 segment 1 [β_3f1; (110-131)]	Ac-AQAIFDDSYLG-NH_2	1743.6
Integrin β_3 segment 2 [β_3f2; (211-220)]	Ac-SVSRNRDAPE-NH_2	1170.5
Glycoprotein IIb fragment	TDVNGDGRHDL	1198.2

Table 3.2 Primary Sequences of studied guest peptides and their masses

Peptide Sequence	Mass (Da.)
Ac-AVTGRGDSPASS-NH_2 (Fibronectin III fragment)	1144.6
Ac-RGDfV-NH_2	633.7
Ac-RGDf[Me]V-NH_2	647.7
c(-RGDf[NMe]V-)	588.6
Ac-RGS(PO3)f[Me]V-NH_2	699.7
Ac-RGDNP-NH_2	598.6
Ac-GRGDNP-NH_2	655.7
c(-Df[Me]VRG-) *Kessler's peptide*	589.5
c(-S(PO4H2)f[Me]VRG-)*Kessler's with Phosphoserine*	640.6
GRGDNP	614.6
GRGDSP	587.6
RGD	343.3
GRGDsP	587.6

Equimolar concentrations of host and guest peptides were analyzed in the ESI-MS to determine which host-guest systems exhibit binding. The tested host-guest systems were obtained through binary combination of the host and guest peptides listed in Tables 3.1 and 3.2. Analyses were carried out using a LCQ Deca XP ESI-ion trap mass spectrometer from Thermo Electron Corporation (West Palm Beach, FL). The mass spectrometer was equipped with a Thermo Surveyor LC-5 Auto-sampler and a Surveyor MS Pump. The ESI source was operated at a spray voltage of 4.5 kV in the positive ionization mode with a N_2 nebulizer flow of 20 arbitrary units. Sample mixtures (20 μL injection) were introduced into the ESI source at a flow rate of 10 μL/min.

For the host-guest pairs that showed binding, system optimization was carried out to enhance the signal of the complex peak formed by the host and the guest peptides. Optimization procedures were carried out (via the auto-tune function in-built in the LCQ instrument) for voltages and manually for capillary temperature and sheath gas flow. Optimization of experimental conditions was also carried out by studying the effects of buffer (ammonium acetate) and acid concentration (acetic acid). The final concentrations of ammonium acetate and acid acetic acid used are 100 mM and 0.5% respectively. The important point borne in mind during all the optimization steps was the enhancement and stability of the complex ion signal. The voltage and temperature conditions were so selected to ensure the preservation of the complex during transfer from the solution phase to the gas phase. Acetic acid was added to assist in the ionization of the species via protonation. The addition of ammonium acetate buffer was necessary to avoid excessive variation in pH during ionization. All experiments were subsequently carried out under the optimized instrument conditions listed in Tables 3.3.

Table 3.3 Optimized instrument conditions for all experiments

Instrument Parameter	Value
Spray Voltage	4.5 kV
Capillary Temperature	200 ^{0}C
Capillary Voltage	30 V
Tube Lens Offset Voltage	15 V
Sheath Gas Flow rate	20 arbs
Auxiliary Gas Flow Rate	0

After the initial screening process and system optimization to enhance the peak of the complex, titration experiments were carried out for each combination of host and guest peptides where appreciable binding was observed, measured by the appearance of a complex peak. For each titration experiment, the concentration of host peptide was held constant at 10 μM while that of the guest was varied between 0.5 μM and 50 μM. Each titration point was measured in triplicate with each measurement taken as the average of 50 sequential scans. An In-house-developed software program was used to calculate dissociation constants for each host-guest binding pair.

Since there existed no literature data for the binding constants for RGD-integrin binding systems studied here, a complementary solution-phase-based capillary electrophoresis binding assay was used to determine the binding affinities of select binding pairs and results compared to

those from ESI-MS as a form of method validation [132] The capillary electrophoresis experiments were carried out in the laboratory of our collaborator (Dr Daniel W. Armstrong, Dept of Chemistry and Biochemistry, UTA). The results of the ESI-MS titration and the capillary electrophoresis experiments are discussed in detail in chapter 4 of this dissertation.

3.3 Experiment 2: Evaluation of the Effects of Physicochemical Parameters on the ESI-MS Response of GXG Tripeptides: An Approach Using Multivariate Statistical Analyses

In this experiment, we explored the possibility of building a statistical model that can be used to correlate analyte ESI-MS response to known or calculable analyte properties. The motivation behind this experiment is the need to be able to correlate the observed analyte signal in ESI-MS to its solution phase concentrations without necessarily making assumptions about its response factor. We have used twelve GXG tripeptides as model analytes to study the effects of their physicochemical properties on their relative ESI-MS response factors. The parameters investigated include solvation energy, nonpolar surface area, polar surface area, nonpolar fractional area, polar fractional area, gas phase basicity, proton affinity, pK_a, and Log D. Each of these parameters is believed to play a role in the equilibrium partition of the analyte between the droplet surface and the inner droplet core.

The tripeptides used in this study were either synthesized (by our collaborator Dr Jung-Mo Ahn, Department of Chemistry, UT-Dallas) or obtained commercially (John Hopkins University Synthesis and Sequencing Facility, Baltimore MD). Three aqueous solvent systems, each containing 100 μM ammonium acetate, were prepared and their pH's adjusted (by adding trifluoroacetic acid, acetic acid, or ammonium hydroxide) to approximately (± 0.1 pH units) 2, 6, and 9, respectively. Solution pH was measured using an Orion pH meter from Thermo-Fisher Scientific (Beverly, MA). Stock solutions (1 mM) of each peptide sample were prepared and 250 μM solutions of each peptide were prepared from the stock solutions by dilution with the appropriate pH-adjusted solvent. Ten percent (10%) methanol was added to each solution to aid

the electrospray process. The response factors were measured under the three pH conditions above. The response factors of the analytes studied herein were determined using a new method developed in our group [112]. This method is based on the Gaussian distribution of analyte concentration that arises as a result of band-broadening dispersion that a plug of analyte undergoes when injected into a solvent flow that is made to pass through a band-broadening element (appropriate length of tubing) before entering the ESI-MS ion source region. The intensity of each analyte was recorded versus time and the data was submitted to a software program developed in-house that calculates the response factor based on a fit to a Gaussian function. This approach alleviates the need for preparing a discrete set of solutions at different concentrations for each analyte to assess concentration dependence of response (i.e. to prepare a calibration curve).

All experiments were carried out using a novel flow injection method (described below) on a LCQ Deca XP ESI-ion trap mass spectrometer from Thermo Electron Corporation (West Palm Beach, FL). Manual injection was used to introduce a small volume (1 μL) of each analyte solution into a solvent flow from an external syringe pump. The syringe pump was operated at a flow rate of 15 μL per minute. An extended length of PEEK tubing (~150μL dead volume) was used to carry the injected sample in the solvent flow from the injection port to the ionization chamber of the mass spectrometer. The ESI source was operated at a spray voltage of 4.5 kV in the positive ionization mode with a N_2 nebulizer gas flow of 20 arbitrary units. The cone voltage and the tube lens voltage were optimized to obtain the highest analyte signal based on the desired molecular ion species. A minimum of five injections were made for each peptide.

3.3.1 Data Collection

The signal intensity of species X is assumed to be directly proportional to its equilibrium concentration in solution:

$$i_X = f_X[X], \quad \text{eq (1)}$$

where f_X is the response factor. When X exists primarily in a singular defined equilibrium state, the equilibrium concentration is the same as its total concentration:

$$i_X = f_X c_X, \quad \text{eq (2)}$$

The concentration of X resulting from flow injection of a known amount of analyte into a band-broadening element (Figure 3.3.1) leading into the ESI source will closely follow a Gaussian distribution over a given solvent volume [112]:

$$c_X = \frac{n_{0,X}}{\sigma_V \sqrt{2\pi}} \exp\left(-\frac{(V - V_p)^2}{2\sigma_V^2} \right), \quad \text{eq (3)}$$

where $n_{0,X}$ is the total amount of X injected, V is volume, V_p is the volume where the peak apex is observed, and σ_V is the standard deviation of the distribution. Because the flow rate (Q) is known and constant, V can be replaced with Qt, V_p with Qt_p, and σ_V with $Q\sigma_t$, where t is time, t_p is the time of peak apex, and σ_t is the standard deviation in units of time:

$$c_X = \frac{n_{0,X}}{Q\sigma_t \sqrt{2\pi}} \exp\left(-\frac{(t - t_p)^2}{2\sigma_t^2} \right), \quad \text{eq (4)}$$

By joining eq. 2 and eq. 4, Equation 5 is obtained:

$$I_X(t) = F_X \frac{n_{0,X}}{Q\sigma_t \sqrt{2\pi}} \exp\left(-\frac{(t - t_p)^2}{2\sigma_t^2} \right), \quad \text{eq (5)}$$

The response factor can then be obtained by fitting a curve into the experimental dependence of signal on time. F_X (henceforth referred to as RF) and σ_t are variables in the fitting; the t_p is defined as the time that divides the area of the peak into halves.

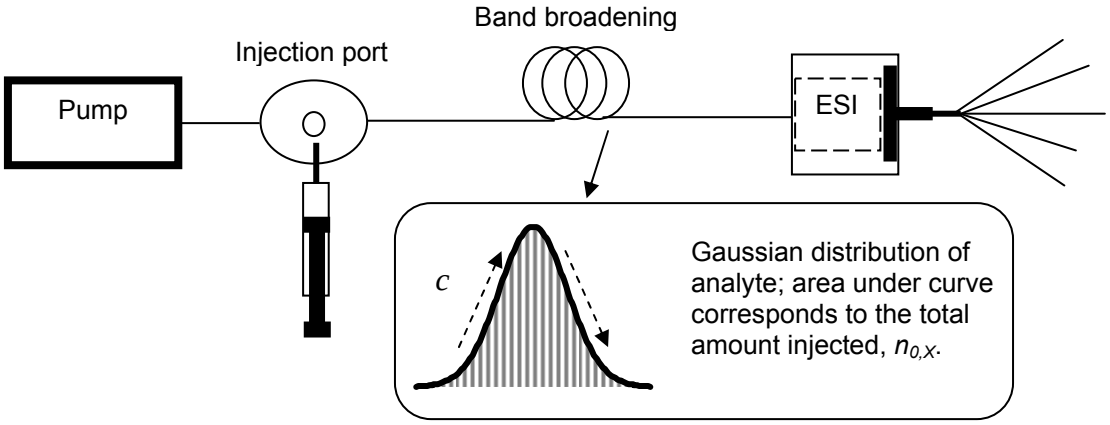

Figure 3.1: Schematic of the flow injection analyte band-broadening method for response factor determination. Shown is an ideal representation of the Gaussian distribution for the analyte after emerging from the band-broadening element (c_X increases on the upslope and decreases on the down slope of the peak; the area under the curve is equal to $n_{0,X}$).

For each peptide, 1 μL volume of 250 μM concentration (2.5 x 10^{-10} mol) was injected into the solvent flow. The injected plug of analyte experiences a band-broadening effect as it disperses in the solvent flow through PEEK tubing (7.5 ft length; 0.10 in i.d.) leading into the ESI source. The resulting Gaussian peak has a width at base of about 2.5 min (σ = 27 s) and constitutes approximately 275 full scans, each of which is a composite of three microscans. In essence, this analysis amounts to a calibration plot incorporating more than 200 data points. The average concentration of the infused peptides at the apex of the band-broadened peak (i.e. the highest concentration) is approximately 20 μM. This is well within the linear dynamic range of response for these peptides when separate calibration plots were performed using discrete solutions of different concentrations (0.5 μM -100 μM) for each peptide (data not shown; limit of linearity is ~ 50 μM).

3.3.2 Physicochemical Parameters

The experimental response factors are correlated with several physicochemical parameters that are either tabulated in the literature or calculable. The parameters are divided into two groups: those referring to individual amino acid (X in GXG) and those referring to the whole GXG molecule. Consideration of parameters pertaining only to X is justified by the use of a model analyte set that varies only by this constituent. The following definitions are given to each parameter as it pertains to this work.

3.3.2.1 Properties for amino acid X in GXG

The following physicochemical properties refer to the middle amino acid in all the GXG tripeptides:

$\Delta G(X)_{g\text{-}w}$ (*Solvation energy*): This is the change in Gibb's free energy (kcal/mol) for the hydration (solvation energy in an aqueous medium) of amino acid X in GXG [133].

GPB (X) (*Gas Phase Basicity* of amino acid X in GXG): This is the negative of the change in Gibbs free energy (kcal/mol) associated with the protonation of the amino acid X in the gas phase [134, 135].

$$X_{(g)} + H^+_{(g)} \longrightarrow HX^+_{(g)} \quad \Delta G = - GPB(X) \qquad \text{Eq. 6}$$

PA (X) (Proton Affinity for amino acid X in GXG): This the negative of the enthalpy (kcal/mol) change associated with the protonation of the amino acid X in the gas phase [134, 135].

$$X_{(g)} + H^+_{(g)} \longrightarrow HX^+_{(g)} \quad \Delta H = - PA(X) \qquad \text{Eq. 7}$$

$pK_a(X)$: (pK_a for amino acid X in GXG): This is defined as -log K_a where K_a (acid dissociation constant) is defined for the solution phase ionization reaction

$$HX_{(aq)} \longrightarrow X^-_{(aq)} + H^+_{(aq)} \qquad \text{Eq. 8}$$

The value of this parameter is calculated using ACD/Labs pK_a Predictor (Ver.9, Advanced Chemistry Development Inc., Toronto, ON) and considered here for the X amino acid in GXG that has an ionizable group in its side chain. For those X amino acids that do not have an ionizable

group on their side chains, an estimate is made from the pK_a of a compound with the closest structure to this side chain (e.g. the pK_a of F (phenylalanine) in GFG was estimated using that of toluene [136]). The two ionizable groups on the flanking G amino acids have similar pK_a values for all the GXG's. The average calculated values for the two pK_a's are 3.57 and 7.63 for the C-terminal acid group and the N-terminal basic group, respectively.

3.3.2.2 Properties for GXG peptides

The following physicochemical properties refer to the entire GXG tripeptides:

*NPSA (Nonpolar surface are*a, $Å^2$): This is the portion of the solvent accessible surface area of the peptide that has overall nonpolar character. The nonpolar surface area of the GXG tripeptides were obtained from tabulated literature data [137].

PSA (Polar surface area, $Å^2$*):* This is the portion of the solvent accessible surface area of the peptide that has overall polar character. The polar surface areas of the GXG tripeptides were obtained from tabulated literature data [137].

TSA (Total Surface Area, $Å^2$*):* This is the sum of NPSA, PSA, and the main chain area (~ 43 $Å^2$ (average)) for all GXG tripeptides [137].

NPFA (Nonpolar Fractional Area): This is the ratio of NPSA to TSA.

PFA (Polar Fractional Area): This is the ratio of PSA to TSA.

Log D (*pH-dependent Log P* (octanol-water partition coefficient, K_{ow})). This parameter is calculated using ACD/Labs Log D calculator.

Charge state (z): This is the charge state of the different species of the GXG's at particular pH values. This parameter is calculated based on calculated pK_a and Log D data using ACD/Labs Log D calculator (Ver. 9, Toronto, ON). The pH conditions under which this study was carried out were selected such that the vast majority of the analyte species exist in a particular charge state.

Average Response Factor: This is the average value of five ESI-MS response factor measurements, as described above, for each peptide at each pH condition using a standard ESI-MS instrument.

Relative RF: This is taken as the ratio of response factor of each peptide to that of GGG, the simplest (structurally and functionally) GXG peptide.

3.3.3 Multivariate Statistical Analyses

Response factors were calculated as described above for each injection, leading to a total of 60 responses (five replicate determinations for each of the 12 GXG tripeptides). With the help of our collaborator (Dr Seoung B. Kim, Dept of Industrial and Manufacturing Engineering, UTA), multivariate statistical models were constructed for each of three separate pH conditions to investigate the relationship between the relative RF and various physicochemical parameters, described in Section 2.5. Constructed models can also be used for making prediction of future unknown samples. In the presented study, multiple linear regression (MLR) models and decision tree models were employed. MLR is a parametric approach that renders a linear equation to examine the relation of the mean response to multiple predictor variables [138]. The coefficient of each predictor variable in the linear equation is estimated by a least squares estimation technique that minimizes the summation of the squared deviation between the actual and fitted values. MLR models have been widely used for prediction problems due to their simplicity [139]. However, MLR models may lead to inefficient and unsatisfactory conclusions when the relationship between the response and predictor variables is nonlinear. Moreover, a parametric assumption of error term in MLR often restricts the applicability of many complicated multivariate data. In order to address this problem, a regression tree model, one of the widely used nonparametric multivariate models, was investigated. Regression tree analysis builds a tree model by splitting the predictor variable space into regions with similar values in the response variable [140]. MATLAB (MathWork Inc., Natick, MA) was used to build the MLR and decision tree models.

3.4 Experiment 3: Chemometric Study of the Effects of Instrumental Parameters on the ESI-MS Response of GXG Tripeptides using Factorial Design

Full factorial experimental design technique was used to study the main effects and the interaction effects in two mass spectrometers equipped with electrospray ion sources. Four major parameters were selected and studied in both instruments, leading to a total of sixteen experiments performed for each run on a single instrument. Significant parameters were identified by plotting the cumulative probability of each treatment against the estimated effects in normal plots. Analysis of variance (ANOVA) was also employed to evaluate the statistical significance of the effects of the parameters on ESI-MS analyte response. The results reveal a number of important interactions in addition to the main effects for each instrument.

The three analytes studied were either synthesized (by Dr Jung-Mo Ahn, UT-Dallas); GRG), or obtained commercially (John Hopkins University Synthesis and Sequencing Facility, Baltimore MD; GDG and GFG). Glacial acetic acid, ammonium acetate, HPLC grade methanol, and LC-MS grade water were all obtained from J.T Baker (Phillipsburg, NJ). The peptides were dissolved in equimolar concentration in a 90/10 water/methanol solvent system containing 1mM ammonium acetate and 0.5% acetic acid. Ten percent (10%) methanol was added to aid the electrospray process.

All experiments were carried out via direct infusion of the analytes using a syringe pump. In the case of the LCQ Deca XP (Thermo Electron Corporation, West Palm Beach, FL), the built-in syringe pump was utilized for the infusion, whereas an external syringe pump was used to achieve the same purpose in the case of the Shimadzu LCMS-2010. Infusion flow rate was kept at 10 μL per minute throughout all the experiments. The parameters investigated were varied according to the design matrix displayed on Table 3.4. In all, four factors were studied at two levels each, leading to an experimental design comprising 16 experimental runs (or treatments), or 2^f (where f is the number of factors evaluated). Each treatment is a unique combination of all the factors.

Table 3.4 Design matrix for the factorial experimental design in random order of experiments

Treatment Number	Spray Voltage	Tube Lens Voltage	Capillary Temperature	Capillary Voltage
1	-	-	-	-
3	+	-	-	-
2	-	+	-	-
6	+	+	-	-
4	-	-	+	-
7	+	-	+	-
9	-	+	+	-
11	+	+	+	-
5	-	-	-	+
8	+	-	-	+
10	-	+	-	+
13	+	+	-	+
15	-	-	+	+
12	+	-	+	+
14	-	+	+	+
16	+	+	+	+

Notes: (-) indicates factor at low level while (+) indicates high level.

Table 3.5: Values of the factors at the two levels investigated

Parameter	Spray Voltage (kV)	Tube Lens Voltage [a] (V)	Capillary Temperature (^{0}C)	Capillary Voltage (V)
Low Level	3 [b]	10	100	10
High Level	6 [b]	70	250	60

[a] The equivalent of this parameter in the Shimadzu LCMS-2010 is Q-array dc voltage.
[b] These values are 2.5 kV and 5 kV respectively for the Shimadzu LCMS-2010.

As shown in Table 3.4, negative signs represent low levels of factors while positive signs represent high levels. The same design matrix was used for all the experiments on both instruments. However, the low and high values differ slightly due to inherent differences in instrument design from different manufacturers. For example, whereas the spray voltage in the Thermo Finnigan LCQ Deca XP can be set as high as 8 kV, the maximum on the Shimadzu LCMS-2010 is 5 kV. Also, the dc component of the Q-array voltage in the Shimadzu LCMS-2010 was taken as the equivalent of the tube lens voltage for the Thermo Finnigan LCQ Deca XP. Experiments were carried out at 25 μM and 50 μM. The experiments at 25 μM were done in triplicate while those at 50 μM were carried out in duplicate.

In order to ascertain whether the rf component of the Q-array voltage in the Shimadzu LCMS-2010 impacts the analyte response, a complementary set of experiments was carried out that incorporates this factor into the treatments; this led to a design of 32 distinct experiments (not shown). This was only carried out on the Shimadzu LCMS-2010 as the LCQ Deca XP does not possess an equivalent parameter. Since no comparison is sought among analyte intensities, no internal standard was needed, and the absolute ion intensities of each analyte were taken directly from the spectra.

The signal intensity of each analyte was taken as the average of about 200 scans in each replicate run per treatment. The average of the signal intensities from all replicates was used as the response for each analyte in each treatment. Once all averages were computed, the

treatments, alongside their respective average responses for each analyte, were rearranged according to Yates' standard order.

Yates' algorithm [141] was used to compute the effects of each treatment by pairing the responses (from top to bottom of the Yates' table) in such a way that no single value occurred in more than one pair. Successive columns were then generated from the column of the response such that the first half of the values in each successive column was a set of values obtained by summing each pair of values in the preceding column and the other half was a set of values obtained by subtracting the bottom value of each pair from the top value. The procedure was repeated to cover the full factorial design. The resulting values in each row of the last column were then divided by the respective divisor. Since there were 16 treatments in all for each experimental run, the final column for the first row was divided by 16 (this treatment is where all parameters are at low values). Each of the other treatments was divided by 8 (since each parameter has half of its treatments at low values and the other half at high values).

The procedure was repeated for the response data acquired on both instruments and at the different concentrations. The final tables of results obtained from Yates' algorithm are presented in chapter 6.

CHAPTER 4

ESI-MS -BASED MOLECULAR RECOGNITION STUDIES BETWEEN INTEGRIN FRAGMENTS AND RGD-BASED PEPTIDES

4.1. Introduction

In this work, titration method was used to evaluate binding affinities between integrin fragments obtained through standard peptide synthesis methods and selected RGD-based peptides. A data treatment method that fits the gas-phase ion abundance data obtained from ESI-MS to a quadratic equation based on 1:1 association between the host and guest peptides was employed. Dissociation constants are then obtained from the fit of this quadratic equation to experimental data. This particular method was first used by Schug et al where it was shown that relative affinity and selectivity values in small-molecule chiral recognition systems correlated well with that determined by microcalorimetric methods and chiral liquid chromatography [105]. Lacking chromatographic or calorimetric data for the RGD-Integrin system studied here, a complementary solution-phase-based capillary electrophoresis binding assay was used to correlate our results.

Capillary electrophoresis is a popular solution phase technique for evaluating molecular associations in many host-guest systems. It has many advantages over other solution phase techniques, including short analysis time and small sample consumption. In addition, there are at least six different experimental approaches that can be explored using CE for quantitative measurement of binding affinities [142]. The capillary electrophoresis experiments were carried out by our collaborators (Dr Daniel W. Armstrong).

Gas-phase collision threshold measurements were also carried out to evaluate the gas phase binding affinity of the integrin fragments and the RGD-based peptides. By varying the collisional excitation in small increments, the voltage required to dissociate 50% of a selected

complex ion was determined as V_{50} [143]. This value measures the stability of the complex in the absence of solvents. As such, relative binding affinities obtained by this method should be viewed with caution when compared to those taken strictly from solution phase methods.

The goal of this work is to extend the use of molecular recognition mass spectrometry techniques to studying biochemically relevant peptide-peptide interactions. This work represents a first approach for studying the viability of assembling specificity of protein-ligand interactions through the analysis of peptide-based contacts in integrin-RGD recognition systems by mass spectrometry. Specific integrin fragment peptides were chosen based on crystallographic and literature data which highlighted the specific binding pockets in integrin subunit proteins [9, 44, 55, 56]. Commercially available and synthetic RGD-based peptides were screened against the integrin fragments to study changes in selectivity and affinity resulting from sequential variation of the RGD peptide ligands. Significant variation is shown when single residues adjacent to the RGD motif are permuted. This fundamental, yet application oriented, approach is expected to be useful in related future high-throughput combinatorial approaches to studying protein-ligand interactions on a smaller scale. Relevant points, including specific advantages and disadvantages are discussed.

4. 2 Results and Discussions

4.2.1 Integrin fragment – RGD peptide binding by ESI-MS

One of the goals of this study is to demonstrate the applicability of ESI-MS as a screening tool for potential peptide-based anti-tumor agents that will target integrins which are over-expressed in tumor cells. The results presented in this dissertation have shown that binding specificity and selectivity between integrin fragments and RGD-containing peptides can be evaluated using ESI-MS. The integrin fragment peptides were chosen because they represent the specific binding pockets in integrin subunit proteins as pointed out by crystallographic and literature data [9, 44, 55, 56]. As shown in Figures 4.1a-c below, the ion abundances

corresponding to the host and the guest are easily distinguishable from each other, as well as from those corresponding to the complex in the ESI-MS spectra; making it possible to evaluate binding affinities by monitoring the intensities of the free host ions and the complex ions at different host-guest concentrations.

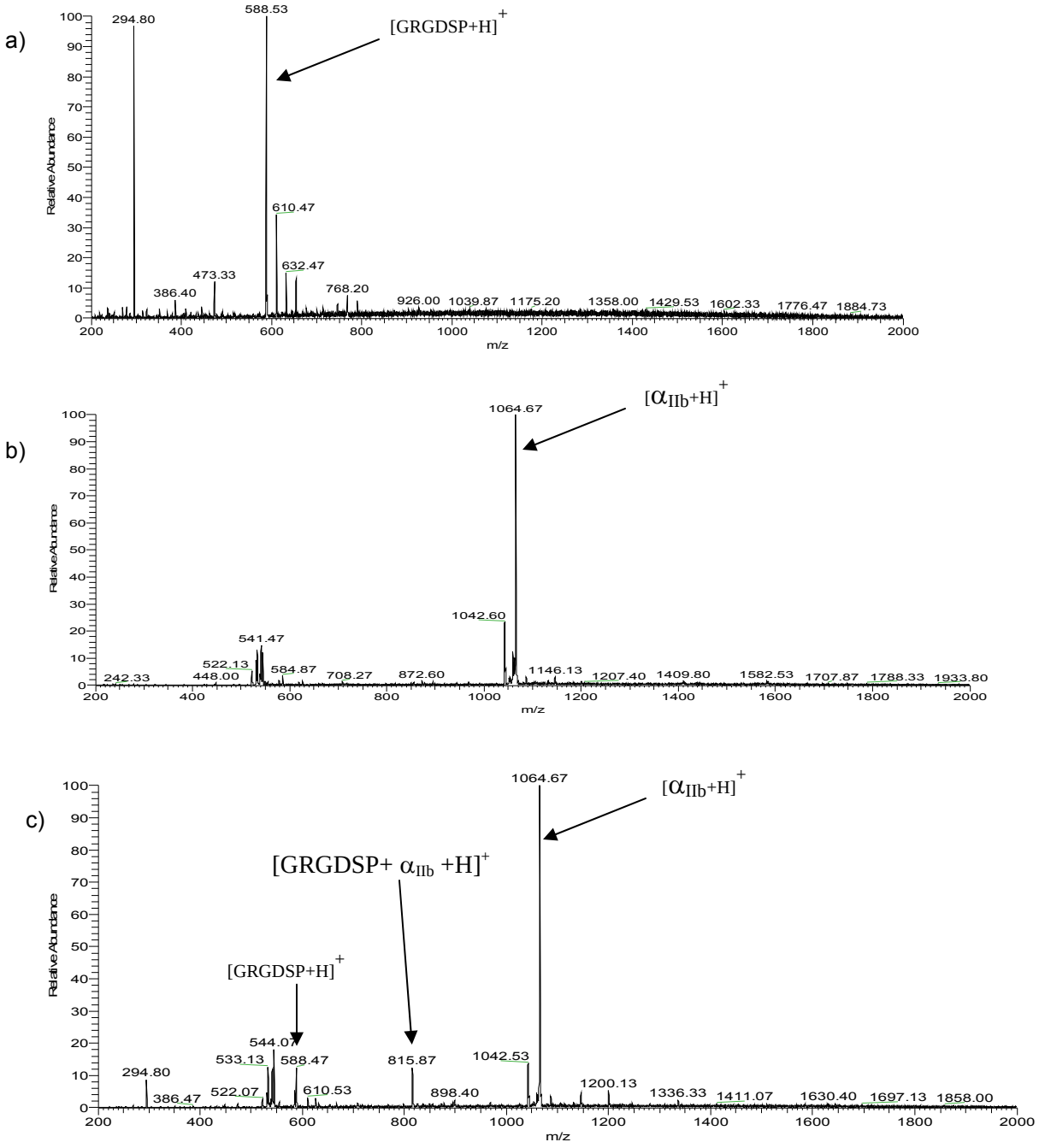

Figure 4.1: ESI mass spectra of (a) GRGDSP at 10μM; (b) α_{IIb}f at 20μM; and (c) mixture of α_{IIb}f and GRGDSP each at 10μM; 50/50 water/methanol solvent condition

4.2.2 Determination of Binding Constants

A titration approach was adopted to evaluate dissociation constants from ion abundance data obtained using ESI-MS. A range of guest peptide concentrations (0.5, 1, 2, 5, 10, 20, and 50μM) were titrated against the host peptide (held fixed at 10μM concentration). The guest concentration range was chosen so as to cover at least two orders of magnitude while taking into consideration the limits of linearity (at high concentration) and limits of detection (at low concentration) for the instrument. The ion abundances measured from these titration experiments can be assumed to relate to the equilibrium concentrations of the different species in solution according to their respective response factors. Given the equilibrium expression for the host (H) - guest (G) complexation as shown in eq 1, we can write an expression for the dissociation constant, K_d (eq 2).

$$H + G \rightleftharpoons HG, \qquad \text{(eq 1)}$$

$$K_d = \frac{[H][G]}{[HG]}, \qquad \text{(eq 2)}$$

This relationship is then used to derive a model for measurement of K_d based on host-guest and host ion abundances.

$$[G] = C_{iG} - [HG], \qquad \text{(eq 3)}$$

where C_{iG} is the initial concentration of guest. Substituting eq 3 into eq 2 and further manipulations yield the new expression for K_d in eq 4.

$$K_d = \frac{[H](C_{iG} - [HG])}{[HG]}, \qquad \text{(eq 4)}$$

If we define degree of association as $\frac{[HG]}{[H]+[HG]}$, then

$$[HG] = C_{iH}\left(\frac{[HG]}{[H]+[HG]}\right), \qquad \text{(eq 5)}$$

$$K_d = \frac{[H]}{[HG]}\left(C_{iG} - \left(C_{iH} \left(\frac{[HG]}{[H]+[HG]} \right) \right) \right), \qquad \text{(eq 6)}$$

We can also define the response factor, which relates the ion intensity of a species to its solution concentration as $f_x = \frac{i_x}{[X]}$; (where [*X*] is the equilibrium concentration of species *x* and i_x is its absolute ion intensity taken from the mass spectrum). Substituting this expression for the different species, followed by subsequent manipulation will yield the expressions,

$$\frac{[H]}{[HG]} = \frac{\frac{i_H}{f_H}}{\frac{i_{HG}}{f_{HG}}} = \frac{i_H}{i_{HG}} \left(\frac{f_{HG}}{f_H} \right) = \mathrm{IF} \qquad \text{(eq 7)}$$

Equation 7 can be substituted in eq 6 to obtain a quadratic equation:

$$(\mathrm{IF})^2 C_{iG} + \mathrm{IF}(C_{iG} - C_{iH} + K_d) - K_d = 0 \ , \qquad \text{(eq 8)}$$

where C_{iH} is the initial concentration of the host and the other variables are defined as above. Equation 8 represents the model algorithm for 1:1 binding to which the ESI-MS experimental data is fit to obtain K_d

The value of "I" can be obtained from the spectrum for each concentration point. If we assume that the response factor of the host and that of the complex are similar, then F can be approximated to unity. Although this assumption may not be strictly valid, there does not yet exist a suitable method whereby the solution phase equilibrium complex concentration can be determined (and thus the response factor of the complex) without first knowing the association constant of the interaction. This point represents perhaps the biggest disadvantage in the use of a MS-based titration approach for studying molecular recognition by small molecules. Even so, this approach has been previously shown to provide a good estimation of relative solution phase binding strengths [105].

By iteratively solving the quadratic equation (eq 8) with various values of K_d in the range 1 to 1000μM and fitting the curves generated therefrom to the experimental data, the K_d value can be obtained as the one that confers the best fit between the experimental curve and the model quadratic curve [105] An example of an ESI-MS titration plot for α_{IIb}f–GRGDNP is shown in Figure 4.2. The titration plot obtained from CE measurements is shown in Figure 4.3 for α_{IIb}f–GRGDSP and the collision threshold measurements for α_{IIb}f-GRGDSP and α_Vf-GRGDSP are shown in Figure 4.4. The K_d values obtained by ESI-MS titration for each of the integrin fragment-RGD peptide host-guest pairs are shown in Table 4.1.

Table 4.1: Comparison of association constants (K) obtained from ESI-MS to those obtained from CE. Also shown is the trend of the threshold collision-induced dissociation measurements.

Host	Ligand	K (ESI-MS) (M^{-1})	K (CE) (M^{-1})	V_{50} (V)
α_{IIb}f	GRGDSP	3800	91	0.935
α_{IIb}f	GRGDNP	6000	67	0.940
α_{IIb}f	GRGDsP	2100	39	0.910
α_vf	GRGDSP	6100	480	0.980
α_vf	GRGDNP	4700	2300	0.985
α_vf	GRGDsP	1500	480	0.960

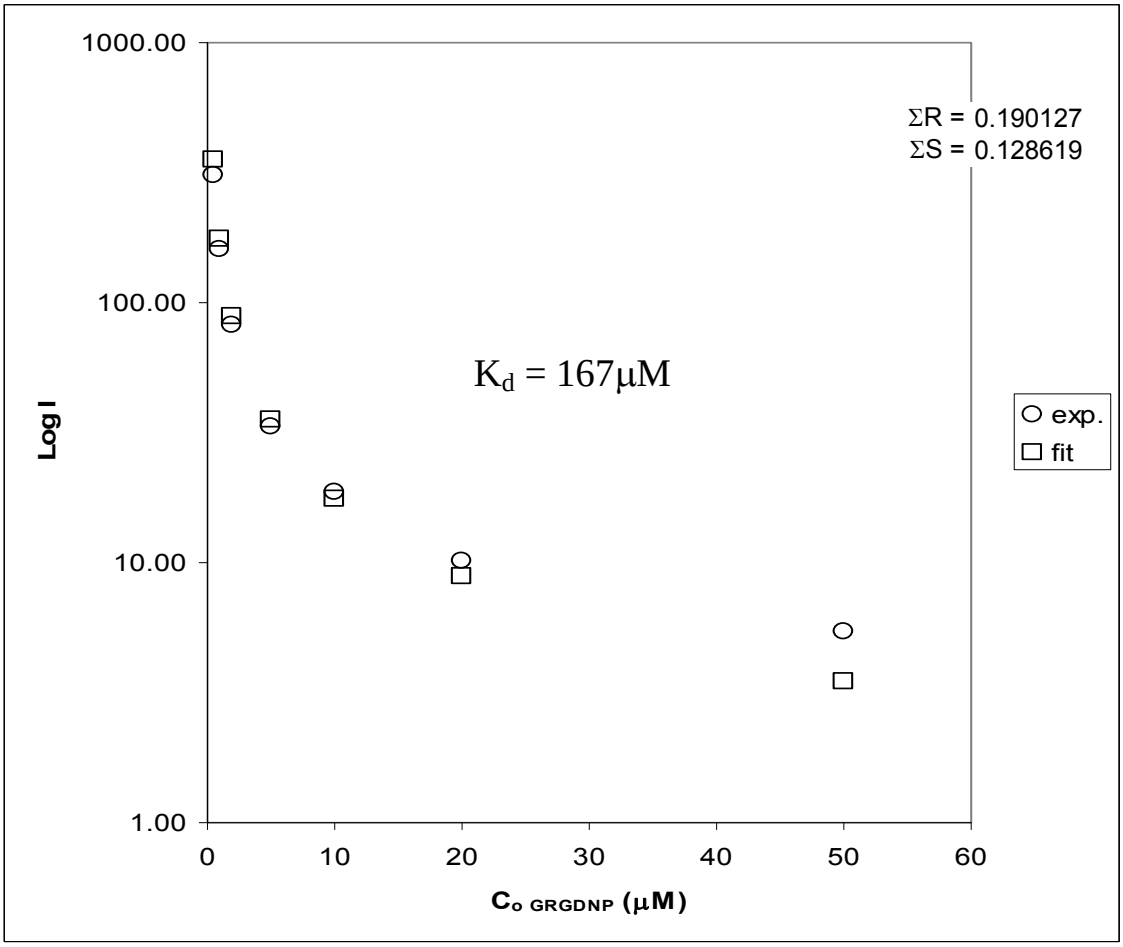

Figure 4.2: ESI-MS Titration plot for α_{IIb}f – GRGDNP. ΣR and ΣS represent, respectively, point-to-point and slope-to-slope weighted difference between the experimental and best model fit data points

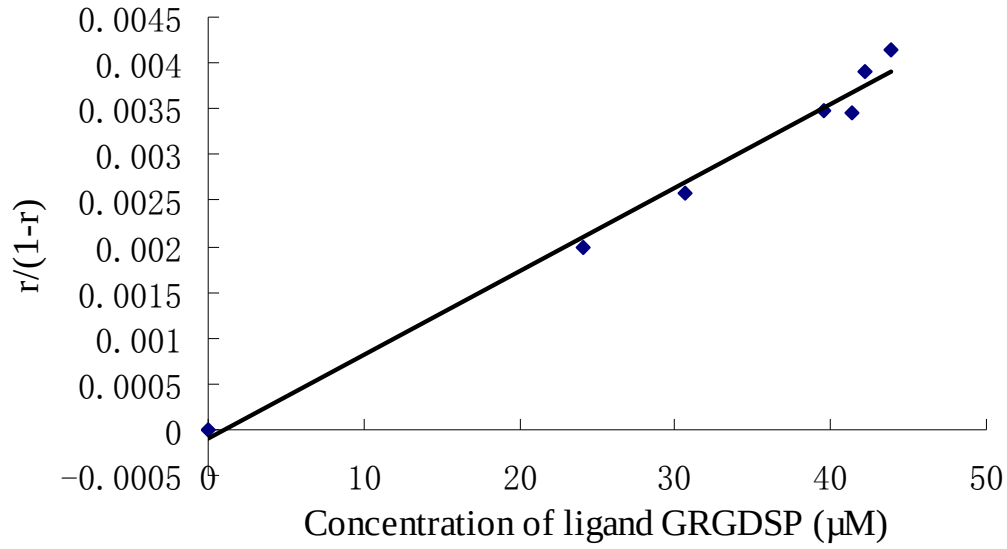

Figure 4.3: CE frontal analysis (FACE) plot for GRGDSP and $\alpha_{IIb}f$ (where r is the fraction of ligand bound per $\alpha_{IIb}f$)

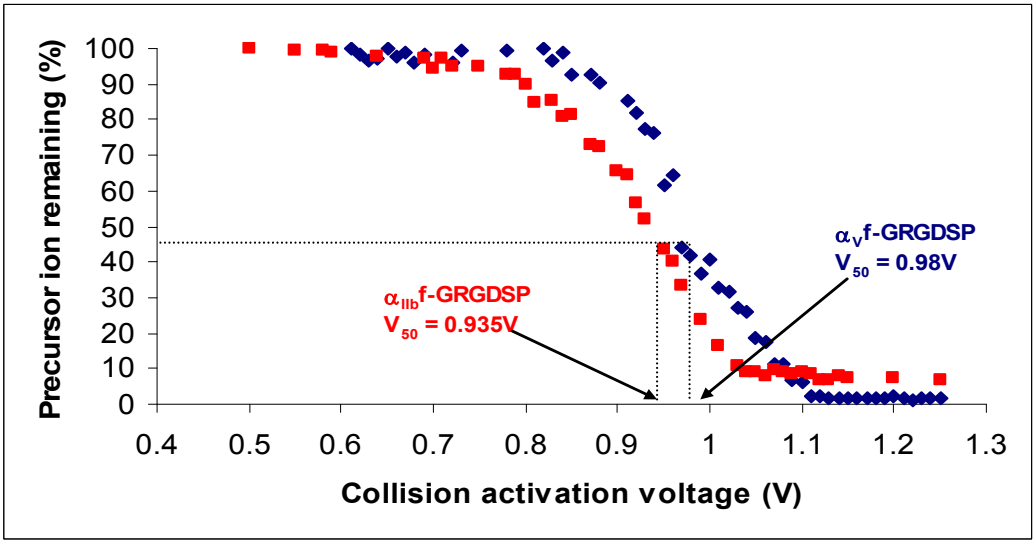

Figure 4.4: Collision threshold measurements for α_{IIb}f-GRGDSP and α_Vf-GRGDSP complexes. V_{50} is the voltage required to fragment 50% of the parent complex ion

4.2.3 Solution Phase Correlation of Binding Constants

The approach described above is the first attempt to delineate integrin protein-ligand specificities using integrin peptide fragments and RGD peptide ligands by ESI-MS analysis. As such, there exists no published data for the binding constants, either in the solution or the gas phase, for this system. In order to validate the approach, the solution phase binding constants for these peptide-peptide interactions were measured using frontal analysis capillary electrophoresis (FACE) [142]. The binding constants obtained from both ESI-MS titration and FACE for the various combinations of integrin fragments and RGD ligands are shown in Table 4.1. The V_{50} values (the voltage required to dissociate 50% of the precursor ion complex after its isolation in the ion trap) are also displayed. The values obtained from ESI-MS titration are in some cases higher than those obtained from capillary electrophoresis, but a similar trend can be drawn for some of the systems studied using both methods. The higher values obtained from ESI-MS titration are not unexpected; and can be explained based on the different factors contributing to the stabilization of the complex in the solution environment and the gas phase.

Three major factors can be responsible for this observation. First, electrostatic forces are strengthened upon transfer of the noncovalent complex from solution phase to gas phase. This is an important factor in the case of integrin-RGD complexes because the complex formation is believed to be driven mainly by electrostatic forces (by virtue of the strong basicity of the guanidine moiety on arginine and the carboxylic acid residue on aspartate) [144]. Upon transfer from a high dielectric to a low dielectric medium, these forces are strengthened. In addition, the complex experiences loss of hydrophobic or solvophobic interactions upon moving from the solution phase to the gas phase. This is due to loss of solvent as a result of evaporation. It increases the degree of motion of the complex because previously locked regions of the complex (due to solvophobic interactions) are now exposed and free to move. This is often referred to as the “entropic contribution” to the overall binding. Thus, the enhancement of electrostatic forces of attraction (“enthalpic contribution”) in the gas phase, coupled with the reduction in hydrophobic

interactions (“entropic contribution”) is partly responsible for the discrepancy between ESI-MS titration and FACE measurements of binding constants. Secondly, as previously mentioned, if the response factors of the complexes, for different permutations of host and guest, do not scale in a similar manner relative to that of the free host, this could introduce further error in the reported binding constants determined by ESI-MS. Thirdly, we must consider that the transfer of species between the solution phase and the gas phase by ESI is a highly dynamic process. As the charged droplets shrink to emit gas phase ions, an increase in concentration of the species contained therein may alter the initial host-guest equilibrium.

4.2.4 Binding Specificity

The results of this work have shown that it is possible to evaluate specific binding between integrin and RGD peptide fragments. Of the five different integrin fragments studied (α_{IIb}f, α_vf, β_3f1, β_3f2, and Glycoprotein IIb), only α_{IIb}f, and α_vf showed specific binding with some of the RGD peptides. For ESI-MS, α_{IIb}f showed greatest binding with GRGDNP compared to the other RGD peptides tested. The order of binding strength observed for the RGD peptides tested with α_{IIb}f is GRGDNP>GRGDSP>GRGDsP. For α_vf, the trend is reversed between GRGDNP and GRGDSP. In the case of FACE measurements, α_{IIb}f showed greater solution-phase binding for GRGDSP than for GRGDNP while α_vf showed greater binding for GRGDNP than for GRGDSP. Both ESI-MS and FACE measurements of binding affinities reveal that GRGDsP has the weakest binding of all three RGD peptides studied.

Threshold CID measurements also support some of the trends observed in the binding constants. For example, the threshold CID measurements show that α_vf binds more tightly to the RGD ligands than α_{IIb}f; which is similar to that observed from solution phase FACE. On the other hand, the trends observed in threshold CID measurements compared to ESI-MS binding constants are similar only in a few cases. The discrepancy between the trends observed in

threshold CID measurements and the ESI-MS binding constants can again be partially attributed to our inability to accurately determine the response factor of the host-guest complex. In principle, positive correlation between such measurements is not necessarily expected since the titration approach measures solution phase association and threshold CID measures gas phase dissociation in the absence of solvation. The fact that the trend of threshold CID measurements correlates better with FACE measurements, compared to ESI-MS titration measurements, highlights the importance of utilizing complementary strategies to evaluate binding in new systems. Such a result indicates that threshold CID may provide an adequate indication of relative binding strength in this system. For further consideration of these systems in a high throughput setup, this is a key point. CID measurements can be performed very quickly and are particularly amenable to a high throughput setting.

The other host peptides (β_3f1,β_3f2, and Glycoprotein IIb) studied did not show significant binding with the RGD peptides studied. Even though β_3 fragments were chosen because they contain putative ligand binding sites, it is possible that the linear RGD peptides studied do not possess the requisite degree of conformational rigidity to form tight complexes with these integrin fragments. Also, these β_3 fragments contain mainly metal ion-directed binding sites and may only be indirectly involved (for conformational purposes) in ligand binding for the intact protein. [55] The RGD-containing segment of human fibronectin protein, as well as the tripeptide RGD ligands studied did not show significant binding with any of the integrin fragments. This observation further demonstrates the specificity of these peptide-peptide interactions.

4.2.5 Evolution of the Titration Method

There is no doubt that ESI-MS-based techniques for studying noncovalent complexes have significant advantages (such as speed, sensitivity, and exact stoichiometric measurements) compared to traditional solution phase approaches. However, the manner in which mass spectrometric data are treated for the purpose of obtaining binding constants raises some

questions. The most commonly asked questions are: 1) how do the response factors of the different species in solution change in the presence of the other species in solution?; and 2) Do the binding constants measured through MS techniques as detection systems correlate well with those measured via other solution phase techniques? In many cases, favorable correlations have been established between ESI-MS-based determinations of binding constants and those obtained from solution phase techniques [107, 117, 145, 146].

Several different approaches abound for calculating values of association or dissociation constants from data acquired using mass spectrometric methods [3]. The titration method employed in this work is both simple and straightforward. Like many other ESI-MS titration methods that involve just the host and the guest and no reference complex, the issues of response factors of the host and the complex (which both appear in the quadratic equation) still have to be considered. The assumption that the response factors of the host and the complex are similar was necessarily made in order to simplify the approach; mainly because experimental determination of response factors for noncovalent complexes is still an elusive task. Until the issue of response factors is properly resolved, K_d values obtained via this method should be used with caution, especially in the absence of rigorously established data from other solution phase techniques for similar systems. From a practical standpoint, the K_d values obtained by MS can most often be used to determine relative binding affinities (selectivities) rather than absolute values. Nonetheless, our response factor assumption may be valid for systems in which the host and the complex are not too different in size (as in the case of a large host binding a small guest), have similar solvation energies, and the complex remains in the same charge state as the host being considered. Subsequent experiments described in this dissertation were directed towards investigating new approaches for evaluating the influence of different parameters (analyte and instrumental) on the ESI-MS response factors of model analytes. Results from such experiments are expected to enhance proper estimation of response factors in the ESI-MS determination of binding affinities between small molecules noncovalent complexes.

CHAPTER 5

EVALUATION OF THE EFFECTS OF PHYSICOCHEMICAL PARAMETERS ON THE ESI-MS RESPONSE OF GXG TRIPEPTIDES: AN APPROACH USING MULTIVARIATE STATISTICAL ANALYSES

5.1. Introduction

Electrospray ionization mass spectrometry (ESI-MS) has evolved over the years as a versatile tool for chemical analysis. Evidence of wide-spread utilization in the area of qualitative characterization of analytes of different types and sizes, including biological, environmental, geological, and pharmaceutical samples can be found in the literature [147-149]. In spite of much success in the area of qualitative applications, quantitative analysis using ESI-MS still poses challenges, mainly due to the fact that it is difficult to predict the ionization efficiency by which different analytes will be released into the gas phase from the electrosprayed droplets of various compositions [150-152]. A correlation coefficient, or "response factor," may be defined that correlates an analyte's observed ion intensity to its solution phase equilibrium concentration. This analytical challenge is even greater when ESI-MS is coupled to a separation technique like high performance liquid chromatography (HPLC) as, in addition to the inherent influence on analyte response of structural changes, there are also effects due to flow rates and changing solution environment (pH and solvent composition) imposed during the chromatographic process.

The usefulness of independent means for determining a correlation parameter is exemplified considering the pharmaceutical analysis of a drug and its metabolite(s) using ESI-MS or LC-ESI-MS. It is difficult to quantitate drugs and their metabolites using these techniques without having carried out standard calibration. Apart from the additional step introduced by the need for calibration with a standard compound, there is a bigger challenge of finding or synthesizing standard compounds for metabolites of new drug candidates. This problem can be

circumvented if it is possible to predict the relative response factor for an analyte of interest under a specific set of solution conditions.

In this part of the dissertation, the potential of building an analyte response model, correlating empirically determined relative ionization efficiencies with known or calculable physicochemical parameters characterizing a model analyte set is investigated. Multiple linear regression and decision tree models are constructed to characterize relationships between the relative response factor (relative to a specified compound, or internal standard) and physicochemical parameters. Equipped with this model, it should be possible to incorporate physicochemical parameters for a related analyte into the model and predict its relative response factor, and hence its concentration in the original solution medium.

The most challenging aspect of this work is the correct identification of the most important parameters that influence an analyte's response during ESI-MS. Despite several research efforts aimed at delineating the many factors that influence ESI-MS ionization efficiency of analytes, a comprehensive model still remains elusive. This is largely due to the wide variability in analyte behavior under different conditions encountered in the ESI process. This unpredictability in analyte behavior can be thought of as arising from the structural make-up of the analyte, which elicits wide ranging properties in different environments.

The ESI-MS response factor of an analyte can be regarded as being made up of two components, analyte-dependent and instrumental factors. In this section, effort is devoted to studying the influence of analyte-dependent factors. Later in chapter 6, the results of using chemometrics for evaluating the effects of instrument parameters on response factors are discussed. The factors influencing the ionization efficiency of an analyte can be categorized into droplet processes and gas phase processes. Droplet processes refer to those processes that take place inside the droplet, and which determine the ease of ion formation, as well as subsequent ejection, of the formed ions into the gas phase. According to the equilibrium partition model [153-155], these include solvation energy, surface activity, and competitive partitioning. Ion

pairing and electrophoretic migration have also been suggested to play a role [156]. It is hypothesized that droplet processes can be modeled by considering relevant analyte structure-related properties, such as hydrophobicity (measurable as Log D), nonpolar surface area, polar surface area, and solvation energy. Gas phase processes are those processes taking place outside the droplet surface after the ions have been ejected to the gas phase, including those processes leading to the acquisition of charge by a neutral molecule upon or just following ejection of the analyte into the gas phase. Relevant gas phase processes, such as gas phase basicity and proton affinity, have been shown to be correlated with analyte response [157].

Understanding the physicochemical factors affecting ESI-MS response of any species is the first step towards developing an effective method for evaluating its response factor. Earlier work in this area includes that of Lieze and colleagues [158], who demonstrated good correlation between the ESI-MS responses of supramolecular complexes and their solvation energies. Sakairi *et al* have also described a good correlation between the natural logarithms of observed ESI-MS intensities of amino acids and the difference between their hydration free energies and their gas phase binding free energies [159]. A correlative relationship between the response factors of some select GGX tripeptides and their nonpolar surface areas has been reported by Enke *et al* [154].

In this work, twelve GXG tripeptides were used as model analytes to study the effects of their physicochemical properties on their relative ESI-MS response factors. In addition to the solvation energy and the nonpolar surface area, the contributions of other parameters such as polar surface area, nonpolar fractional area, polar fractional area, gas phase basicity, proton affinity, pK_a, and Log D have also been evaluated. The response factors of the twelve tripeptides have been measured under three different pH conditions. Multivariate statistical analysis was used to evaluate the inter-correlation of the different parameters and their correlation with response factors. The parameters that are most correlated with the relative response factors (determined by normalizing the absolute response factors of the tripeptides to the response factor

of GGG) are included in the model for each pH condition. To the best of our knowledge, there has been no earlier account of a predictive model that specifically takes into account the pH-dependent analyte behavior by incorporating such parameter as Log D of an analyte.

The response factors of the analytes studied herein are determined using a new method developed in our group [112]. This method is based on the Gaussian distribution of analyte concentration that arises as a result of band-broadening dispersion that a plug of analyte undergoes when injected into a solvent flow that is made to pass through a band-broadening element (appropriate length of tubing) before entering the ESI-MS ion source region. The intensity of each analyte was recorded versus time and the data was submitted to a software program developed in-house that calculates the response factor based on a fit to a Gaussian function. This approach alleviates the need for preparing a discrete set of solutions at different concentrations for each analyte to assess concentration dependence of response.

5.2 Results and Discussions

5.2.1 GXG Response Factors vs. Physicochemical Parameters – Qualitative Treatment

The objective of this work was to probe the dependence of ESI-MS analyte response on calculable or experimentally determinable physicochemical parameters. According to the equilibrium partition model, it is postulated [153] that the distribution of analyte species between the core of the droplet and its surface (specifically the concentration of the excess surface charge) will determine the overall amount of analyte ions that are released into the gas phase for subsequent mass analysis. This distribution is influenced by factors such as analyte solvation energy, analyte surface activity, competitive partitioning, ion pairing, and electrophoretic migration (droplet processes). These droplet processes can be modeled by considering different analyte physicochemical properties. For example, surface activity is modeled by hydrophobicity (measurable as log D) and nonpolar surface area; and competitive partitioning is highly inter-correlated with the relative surface activity and solvation energy. In addition to the surface/bulk

solvent distribution of analyte ions and neutrals in the droplet, the totality of the ion current observed is also affected by gas phase processes, which are characterized by analyte properties such as gas phase basicity and proton affinity [153, 154]. The calculated response factors, the relative response factors as well as the values of the parameters for the GXG tripeptides studied are tabulated in Tables 5.1- 5.3.

Table 5.1: Values of studied parameters and calculated ESI-MS response factors (pH 2)

GXG Species	z	NPSA ($Å^2$)	PSA ($Å^2$)	TSA ($Å^2$)	NPFA	PFA	$\Delta G(X)_{g-w}$ kcal/mol	GPB(X) kcal/mol	PA(X) kcal/mol	pK_{aCOOH}	pK_a (X)	pK_{aNH2}	Log D	Average RF	RF Relative to GGG
GSG	1	44	36	122	0.361	0.295	-4.31	207.6	215.2	3.55±0.10	14.81±0.10	7.60±0.29	-5.34	5.2E+11	2.21
GLG	1	137	0	180	0.761	0.00	5.20	209.6	217.4	3.59±0.10	51.00	7.66±0.29	-3.35	8.7E+11	3.66
GKG	2	119	48	211	0.564	0.227	-6.84	221.8	235.6	3.58±0.10	10.44±0.10	7.64±0.29	-6.12	3.3E+11	1.40
GHG	2	102	49	194	0.526	0.253	-8.25	223.7	231.5	3.54±0.10	6.70±0.61	7.61±0.29	-6.29	6.5E+11	2.73
GFG	1	175	0	218	0.803	0	2.18	212.1	219.9	3.59±0.10	43.00	7.66±0.10	-2.92	1.2E+12	4.91
GGG	1	0	0	85	0	0	2.42	202.7	210.5	3.58±0.10	17.10	7.66±0.29	-5.11	2.4E+11	1.00
GAG	1	67	0	113	0.593	0	2.63	206.4	214.2	3.59±0.10	48.00	7.66±0.29	-4.76	3.8E+11	1.61
GVG	1	117	0	160	0.731	0	4.07	208.7	216.5	3.59±0.10	51.00	7.65±0.29	-3.88	3.1E+11	1.31
GRG	2	89	107	241	0.369	0.444	-17.46	237.0	244.8	3.58±0.10	13.35±0.70	7.65±0.29	-6.87	2.1E+11	0.89
GYG	1	144	43	229	0.629	0.188	-3.17	213.1	220.9	3.58±0.10	9.84±0.15	7.64±0.29	-3.66	2.1E+11	0.90
GDG	1	48	58	151	0.318	0.384	-9.64	208.6	216.4	3.53±0.10	4.10±0.10	7.64±0.29	-4.63	1.0E+12	4.21
GNG	1	44	69	158	0.278	0.437	-8.31	212.8	220.6	3.57±0.10	13.45±0.46	7.62±0.29	-5.6	7.1E+11	3.00

Table 5.2: Values of studied parameters and calculated ESI-MS response factors (pH 6)

GXG Species	z	NPSA ($Å^2$)	PSA ($Å^2$)	TSA ($Å^2$)	NPFA	PFA	$\Delta G(X)_{g-w}$ kcal/mol	GPB(X) kcal/mol	PA(X) kcal/mol	pK_{aCOOH}	pK_a (X)	pK_{aNH2}	Log D	Average RF	RF Relative to GGG
GSG	0	44	36	122	0.361	0.295	-4.31	207.6	215.2	3.55±0.10	14.81±0.10	7.60±0.29	-4.77	1.13E+12	3.11
GLG	0	137	0	180	0.761	0.00	5.20	209.6	217.4	3.59±0.10	51.00	7.66±0.29	-2.78	4.56E+12	12.56
GKG	1	119	48	211	0.564	0.227	-6.84	221.8	235.6	3.58±0.10	10.44±0.10	7.64±0.29	-5.47	3.04E+12	8.39
GHG	1	102	49	194	0.526	0.253	-8.25	223.7	231.5	3.54±0.10	6.70±0.61	7.61±0.29	-5.42	4.48E+12	12.35
GFG	0	175	0	218	0.803	0.00	2.18	212.1	219.9	3.59±0.10	43.00	7.66±0.10	-2.35	6.53E+12	18.00
GGG	0	0	0	85	0	0	2.42	202.7	210.5	3.58±0.10	17.10	7.66±0.29	-4.54	3.63E+11	1.00
GAG	0	67	0	113	0.593	0	2.63	206.4	214.2	3.59±0.10	48.00	7.66±0.29	-4.19	5.73E+11	1.58
GVG	0	117	0	160	0.731	0	4.07	208.7	216.5	3.59±0.10	51.00	7.65±0.29	-3.31	1.66E+12	4.57
GRG	1	89	107	241	0.369	0.444	-17.46	237.0	244.8	3.58±0.10	13.35±0.70	7.65±0.29	-6.22	2.33E+12	6.43
GYG	0	144	43	229	0.629	0.188	-3.17	213.1	220.9	3.58±0.10	9.84±0.15	7.64±0.29	-3.09	2.30E+12	6.35
GDG	-1	48	58	151	0.318	0.384	-9.64	208.6	216.4	3.53±0.10	4.10±0.10	7.64±0.29	-5.01	2.29E+12	6.30
GNG	0	44	69	158	0.278	0.437	-8.31	212.8	220.6	3.57±0.10	13.45±0.46	7.62±0.29	-5.03	2.75E+12	7.59

Table 5.3: Values of studied parameters and calculated ESI-MS response factors (pH 9)

GXG Species	z	NPSA ($Å^2$)	PSA ($Å^2$)	TSA ($Å^2$)	NPFA	PFA	$\Delta G(X)_{g-w}$ kcal/mol	GPB(X) kcal/mol	PA(X) kcal/mol	pK_{aCOOH}	pK_a (X)	pK_{aNH2}	Log D	Average RF	RF Relative to GGG
GSG	-1	44	36	122	0.361	0.295	-4.31	207.6	215.2	3.55±0.10	14.81±0.10	7.60±0.29	-5.8	8.63E+11	3.18
GLG	-1	137	0	180	0.761	0.00	5.20	209.6	217.4	3.59±0.10	51.00	7.66±0.29	-3.79	3.29E+12	12.11
GKG	0	119	48	211	0.564	0.227	-6.84	221.8	235.6	3.58±0.10	10.44±0.10	7.64±0.29	-4.58	4.56E+12	16.79
GHG	-1	102	49	194	0.526	0.253	-8.25	223.7	231.5	3.54±0.10	6.70±0.61	7.61±0.29	-5.71	5.62E+12	20.71
GFG	-1	175	0	218	0.803	0	2.18	212.1	219.9	3.59±0.10	43.00	7.66±0.10	-3.36	4.92E+12	18.14
GGG	-1	0	0	85	0	0	2.42	202.7	210.5	3.58±0.10	17.10	7.66±0.29	-5.55	2.71E+11	1.00
GAG	-1	67	0	113	0.593	0	2.63	206.4	214.2	3.59±0.10	48.00	7.66±0.29	-5.2	5.62E+11	2.07
GVG	-1	117	0	160	0.731	0	4.07	208.7	216.5	3.59±0.10	51.00	7.65±0.29	-4.32	1.31E+12	4.81
GRG	0	89	107	241	0.369	0.444	-17.46	237.0	244.8	3.58±0.10	13.35±0.70	7.65±0.29	-5.32	2.16E+12	7.97
GYG	-1	144	43	229	0.629	0.188	-3.17	213.1	220.9	3.58±0.10	9.84±0.15	7.64±0.29	-4.15	1.92E+12	7.09
GDG	-2	48	58	151	0.318	0.384	-9.64	208.6	216.4	3.53±0.10	4.10±0.10	7.64±0.29	-6.09	3.15E+12	11.62
GNG	-1	44	69	158	0.278	0.437	-8.31	212.8	220.6	3.57±0.10	13.45±0.46	7.62±0.29	-6.05	5.12E+12	18.86

Attempt to correlate GXG relative response factors with individual physicochemical parameters reveals some interesting patterns. The plot of the variation of response factors with some of the parameters (total surface area, nonpolar surface area (NPSA), GPB, and Log D) are depicted in Figure 5.1 for pH 6 and pH 9.

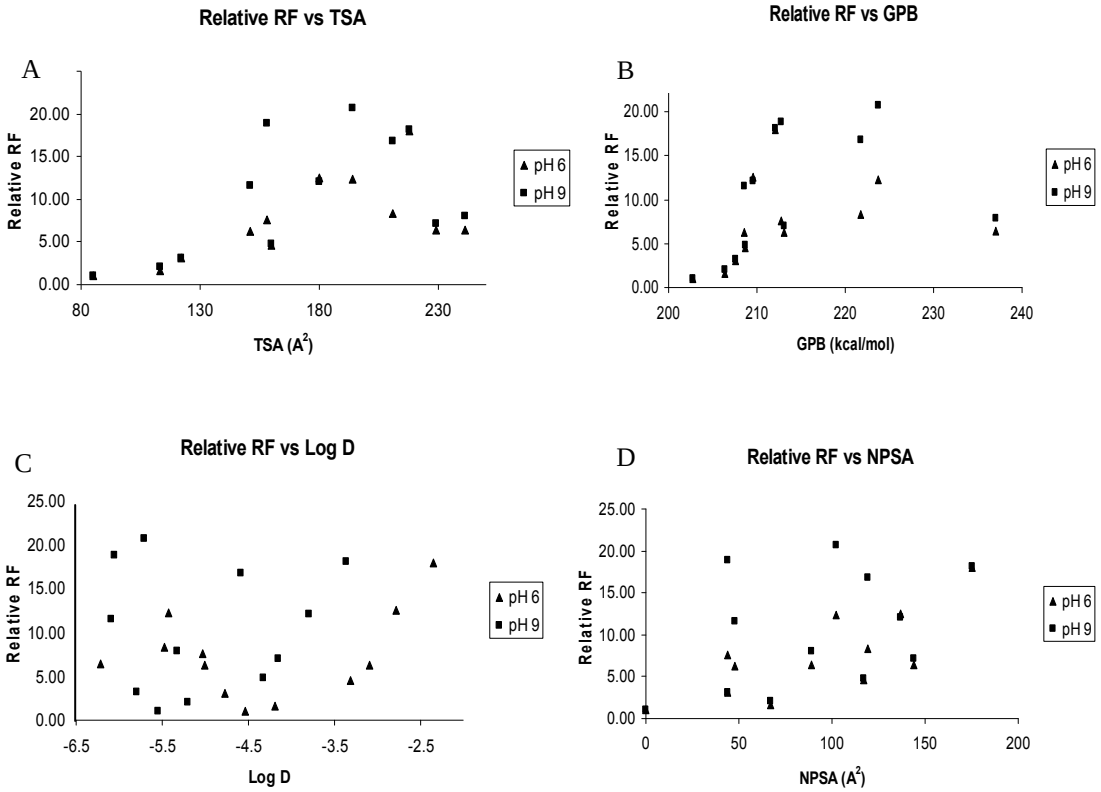

Figure 5.1: Relative RF vs (A) total surface area (TSA), (B) gas phase basicity (GPB), (C) Log D, and (D) nonpolar surface area (NPSA).

Similar plots were drawn for the variation of relative response factors with polar surface area (not shown). Examining the plots in Figure 5.1, it is clear that no single parameter has a discernable linear correlation with relative response factor. From the raw data, it is concluded that total surface area shows the most reasonable single parameter correlation with relative response factor when compared with polar and nonpolar surface areas. Even though the presence of a polar region on the surface of the molecule is assumed to assist in charge acquisition, either in

solution or in the gas phase, polar surface area cannot explain the observed trend of response factors by itself. Likewise, the analytes with higher nonpolar surface areas have greater propensities to become more surface active and therefore eject out of the droplet surface more efficiently than those with less nonpolar surface area. However, similar to the case of polar surface area, nonpolar surface area cannot single-handedly explain the observed trends in analyte response. With the exception of a few cases, higher total surface area translated to higher response factor (Figure 5.1A). Even so, a prediction model cannot be built on knowledge of this parameter alone. It should be further noted that, as this model is extended to different analytes, for which these surface area values are not tabulated, they may be calculated using an available freeware software program such as SurfRacer [160]. For GPB (Figure 5.1B), better correlation with response factor was observed at pH 9 than at pH 6.

The role of hydrophobicity in determining surface activity of an analyte is of key importance. Species with more hydrophobic character tend to migrate more quickly to the droplet surface. The hydrophobicities of the tripeptides were assessed through their calculated LogD and tabulated NPSA properties. As mentioned earlier, ESI-MS ion abundance is assumed to be dependent on the partitioning of analyte ions and neutrals between the core of the droplet and the droplet surface. Based on the fact that Log D is a measure of the distribution of species across an octanol-water phase boundary, it is reasonable to approximate the solution phase (high dielectric constant)/gas phase (low dielectric constant) boundary characteristics of the droplet with that of water/octanol. As such, a large Log D value, indicating higher surface activity, is expected to translate to greater ion abundance. Although it is one of the key factors that are expected to have greater influence on an analyte's ESI-MS response factor, from a qualitative point of view, the Log D property was found to be poorly linearly correlated with ESI-MS response factors (Figure 5.1C).

Investigating further and considering the plots in Figure 5.2, little correlation could be drawn between the relative response factors for the GXG tripeptides studied here and solvation

energies of the X amino acid (ΔG(X)), either alone or in summation with their gas phase basicities, an approach that was successfully demonstrated for single amino acids by Sakairi *et al* [159].

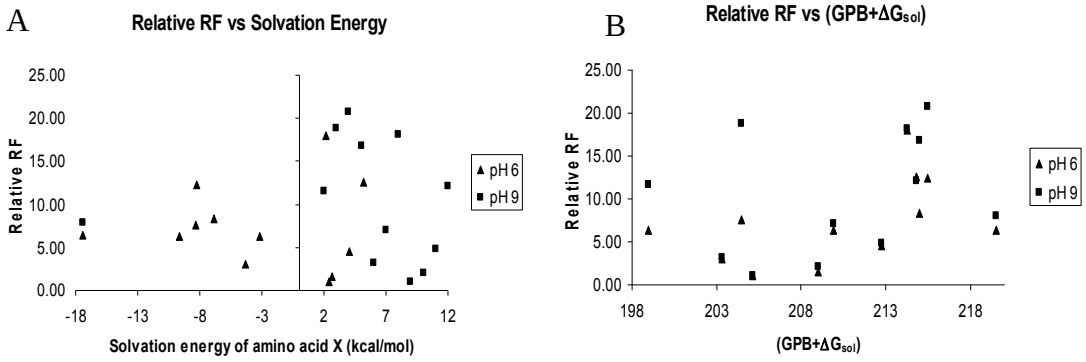

Figure 5.2: Relative RF vs solvation energy (A), and sum of GPB and solvation energy (B)

This is more or less antithetical considering earlier evidence supporting possible correlation between ESI-MS analyte response and analyte solvation energy. For example, in a previous study on the effects of solvation energy on ESI-MS analyte response factors, Lieze *et al* [158] had shown that a simple mathematical relationship could be applied to describe the behavior of alkali metal ions in ESI. In that study, a group of alkali metal chlorides were studied in pure water with all analytes present at the same concentration. It was observed that despite the equimolar concentrations of the analytes, the observed ion signals varied markedly in intensity. Since all measurements were carried out under the same condition, the difference in ion intensity was attributed to the differences in their solvation energies. Sakairi *et al* [159] also reported good correlation between the natural logarithm of observed ESI-MS intensities of amino acids and the sum of their hydration free energies with their gas phase basicity free energies. This observation is remarkably true when dealing with simple analytes like single ions and molecules with relatively simple structure. However, with more functionally and structurally diverse analytes, it is not

surprising that little correlation is observed between analyte solvation energies and observed ion yields.

Even when the peptides with ionizable side chains were considered separately from those with non-ionizable side chains, little correlation with any single parameter was observed. One notable observation is that for the most part, GXG peptides with nonpolar side chains demonstrated better linear correlation between their relative response factors and the parameters depicted in Figure 5.1. It was also noted that these peptides show smaller variability between relative response factors measured at pH 6 and those measured at pH 9 when compared with their counterparts with polar ionizable side chains. These observations suggest that there is indeed an added complexity introduced for analytes with ionizable groups which limits the direct comparison between observed ESI-MS responses of such analytes with those having non-ionizable side chains. For this reason, and for the purpose of quantitative statistical evaluation (*vide infra*), we have excluded some of the parameters that pertain only to the middle amino acid X (G<u>X</u>G), particularly those that play greater role in droplet processes, in an effort to better generalize the model for future applications to a more diverse analyte set.

5.2.2 GXG Response Factors vs. Physicochemical Parameters – Quantitative Statistical Treatment

MLR and decision tree models were constructed to predict relative RF based on the physicochemical parameters. After careful evaluation of the relative contribution of each parameter to the relative RF, alongside consideration for any occurrence of strong correlation between any pair of physicochemical parameters, we decided to use the following five parameters to build the models: PA(X), NPSA, PSA, TSA, and Log D. Most of the parameters pertaining to amino acid X (in GXG) were left out deliberately because we recognize that including these parameters may constitute a limitation regarding the general applicability of the procedure for future elaboration. Also, derived parameters (such as non-polar fractional area = NPSA/TSA) were omitted as they show strong correlation with the parameters from which they were derived.

MRL analyses were conducted using a stepwise regression algorithm based on AIC criteria in the R package (www.r-project.org). The stepwise regression procedure identified the model of each pH that includes the physicochemical parameters shown in Table 5.4.

Table 5.4: Results of stepwise regression analysis in pH 2, pH 6, and pH 9

	pH 2		pH 6		pH 9	
	Coefficients	p-value	Coefficients	p-value	Coefficients	p-value
Intercept	7.89	<0.01	144.03	0.01	-44.06	<0.01
NPSA	0.03	0.07	0.16	<0.01	NA	NA
PSA	0.05	0.02	NA	NA	-0.20	<0.01
Total Area	-0.04	0.06	0.10	0.04	0.18	<0.01
PAX	NA	NA	-0.92	<0.01	NA	NA
LogD	0.68	<0.01	-8.80	<0.01	-5.44	<0.01

Coefficients and p-values of the parameters, not selected by a stepwise regression procedure are indicated as NA (Table 5.4). A small p-values (<0.01) indicates a highly statistically significant parameters in the regression models. Note that the parameter, LogD is highly statistically significant in all three pH models, as indicated by the small p-values, despite the little or no correlation observed during qualitative treatment.

Regression tree analyses were conducted using the R package. Three regression tree models for pH 2, pH 6, and pH 9 were developed for predicting the relative RF based on the five physicochemical parameters of interest (Figure 5.3, Figure 5.4, and Figure 5.5). Each node of the regression tree specifies conditions that split an existing region. For example, in Figure 5.3, the top node splits to the left using the condition that LogD $\leq$ -3.505, and splits to the right using the condition that LogD > -3.505. The final regions are specified by the terminal nodes of the tree. We can determine the criteria for each terminal node by backtracking up the tree to the top node. In Figure 5.3, the first terminal node (leftmost terminal node) backtracks up the left edge of the tree, yielding the following rule: "if LogD was less than -3.505, PSA was less than 48.5 square angstrom, LogD was less than -5.225, and NPSA was less than 81.5 square angstrom, the average of GXG relative RF is 2.214". The first condition (LogD $\leq$ -3.505) is actually redundant because the third condition (LogD $\leq$ -5.225) includes the first condition. Similar interpretations can be made in other terminal nodes in the regression tree models.

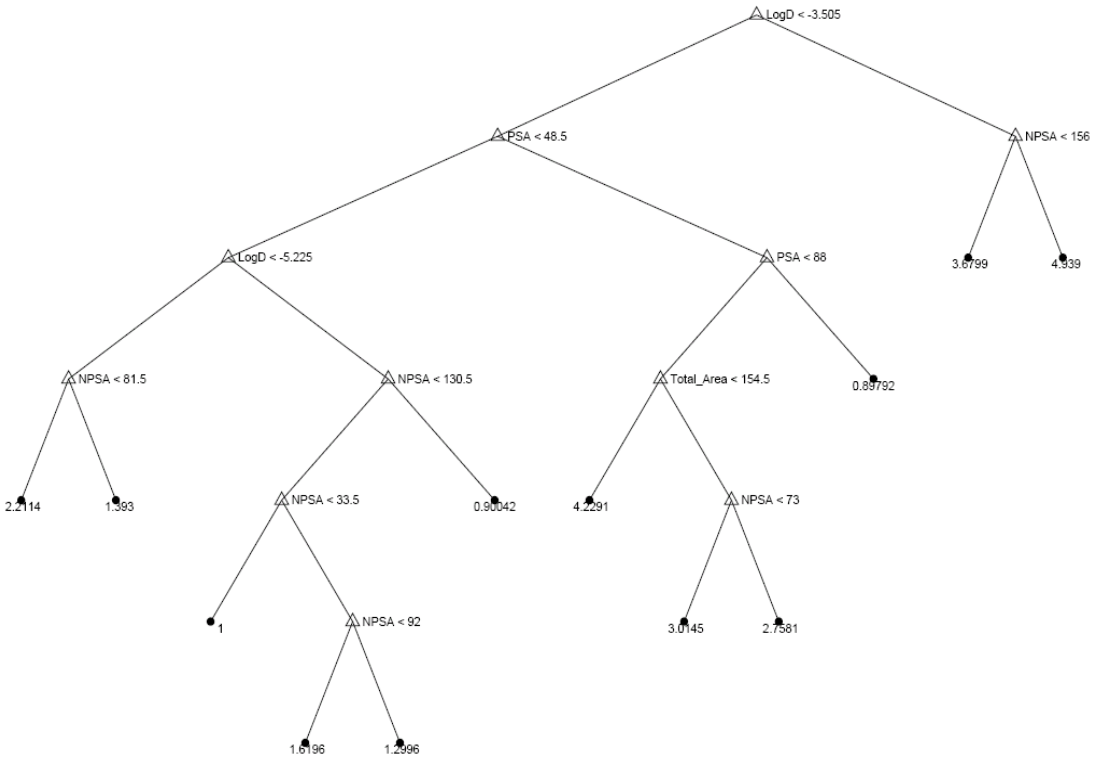

Figure 5.3: Regression tree output for pH 2.

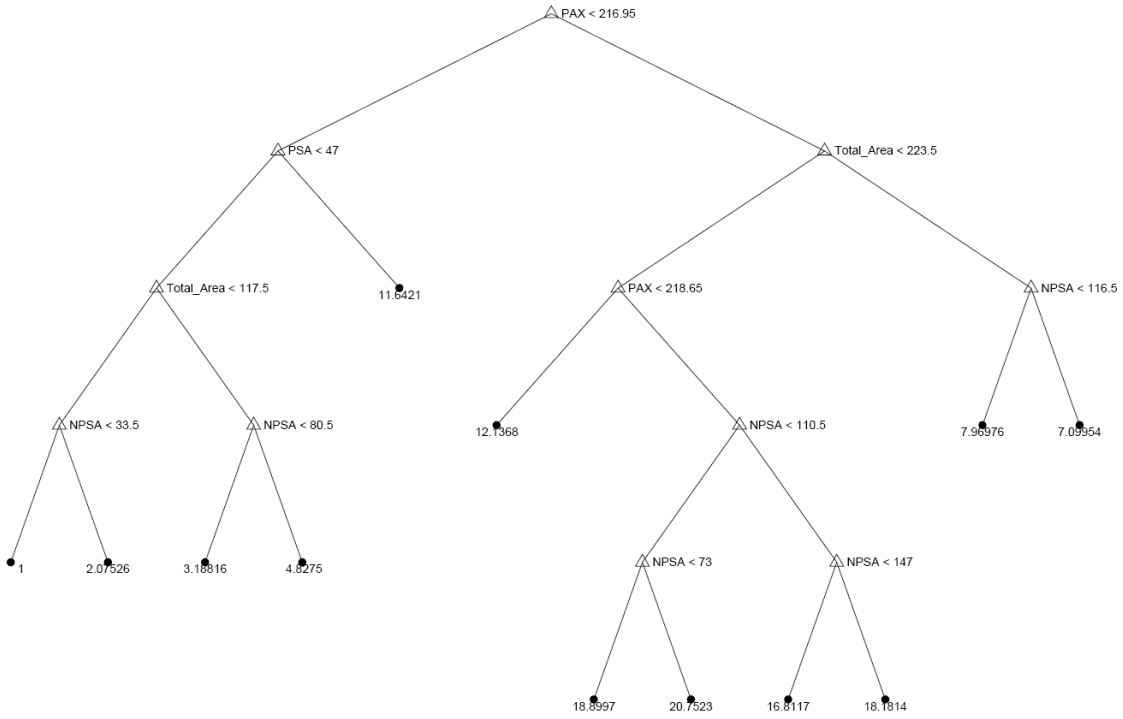

Figure 5.4: Regression tree output for pH 6.

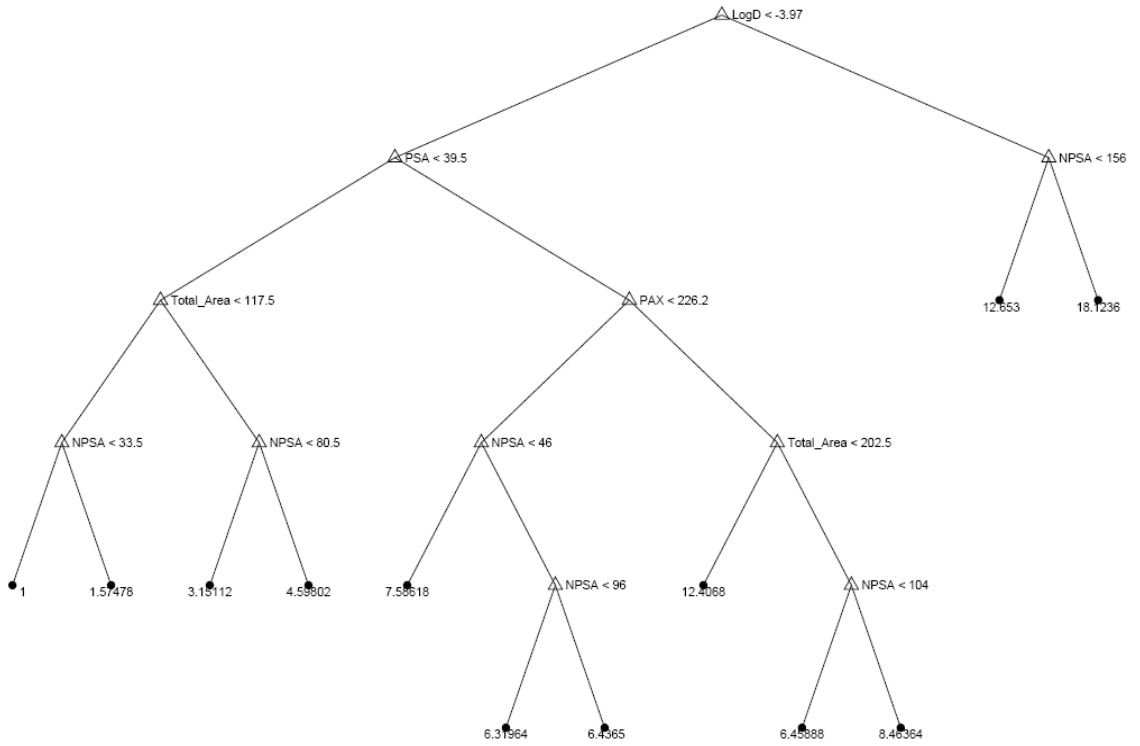

Figure 5.5: Regression tree output for pH 9.

Regression tree models also provide the importance of parameters to predict the response variable. In general, the parameter selected at the top node is the most important one for predicting the response variable and the parameters selected in the second level are the next important, and so on. Consequently, the most important physicochemical parameter to predict the relative RF in the pH 2 model is logD, followed by PSA, NPSA, and Total Area. This result is in general agreement with the regression analysis. In pH 6, both the regression tree and stepwise regression procedures selected PA(X) as the important parameter to predict relative RF. However, logD was selected as the important parameter in the stepwise regression model but not in the regression tree model. Another parameter that shows discrepancy between MRL and regression tree is PSA that was selected as the important parameter in the regression tree model but not in the stepwise regression model. In pH 9, regression tree used all five parameters in its

model construction, while the stepwise procedure in MRL selected only three parameters: PSA, Total Area, and LogD (Table 5.4).

To evaluate the efficacy in predicting the relative RF in the MLR and regression tree models and to compare the two models, a 12-fold cross-validation (CV) was conducted. Specifically, the experimental data was split into 12 groups corresponding 12 different species. In each of 12 rounds in CV, 11 groups were used to train (construct) the model, and the remaining one group was used for testing (validating) the constructed model and calculating the sum of square for error (SSE). SSE is computed by the summation of differences between the actual and fitted values. This process was repeated 11 times more with different sets of training and testing groups to obtain 12 SSE values. The overall SSE or CV-SSE is then the average of these 12 SSE values. Figure 5.6 shows the values of CV-SSE of the MLR and regression tree models in each of the three pH conditions.

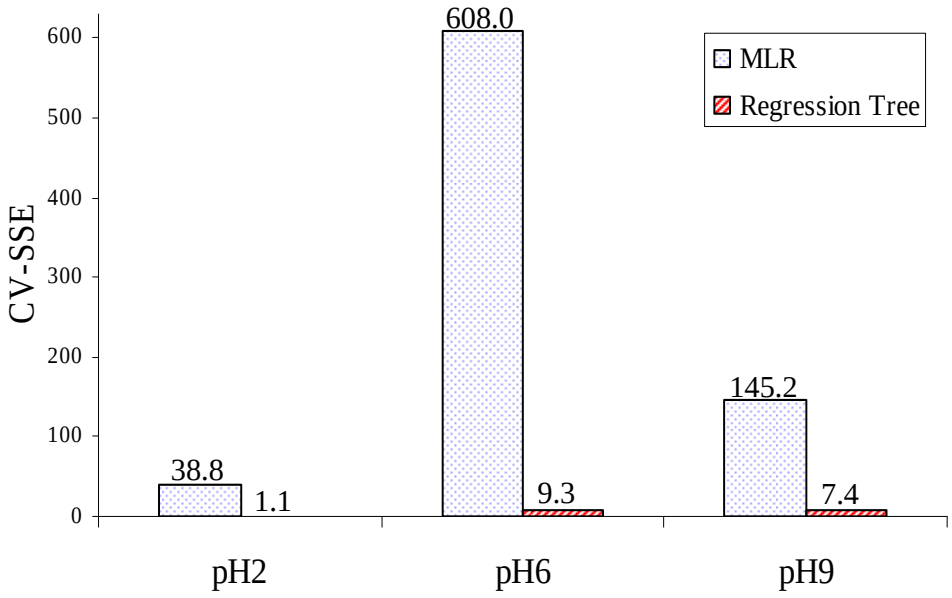

Figure 5.6: CV-SSE of the MLR and regression tree models in pH 2, pH 6, and pH 9.

The results showed that the decision tree models yielded significantly smaller SSE than the MLR models in all three pH levels. This implies that the capability of handling nonlinear structure in a regression tree results in a significant reduction of error in the prediction of relative RF. In particular, the large difference in the magnitude of SSE between MRL and regression tree in pH 6 and pH 9 suggests that the parameters selected by regression tree may be more relevant than those selected by the stepwise regression in MRL. Note that MRL and regression tree use different parameters in the model construction in pH 6 and pH 9. Finally, the result indicates that both the MLR and decision tree models produced the highest SSE in pH 6, while the models produced the least SSE in pH 2. It should be noted that the smaller SSE observed for pH 2 compared to pH's 6 and 9 does not indicate better correlation between response factors and the selected parameters under this condition. To the contrary, response factors measured at pH 2 were generally lower than those measured at pH's 6 and 9 and showed higher noise levels in the spectra compared to the other two conditions. It was more difficult to fit the data at this condition to our computer program in calculating the response factors. The low SSE may have resulted from the fact that the absolute values of the response factors at this pH were low and the impact of the high noise level may have masked any variability that may be attributable to differences in the parameters. On the other hand, spectra measured at pH 9 showed lower noise level than those at pH 6; and the data fit better to the computer program. The observed lower SSE values for pH 9 compared to pH 6 for both the MLR and decision tree models are thus in line with experimental observation.

CHAPTER 6

CHEMOMETRICS STUDY OF THE INFLUENCE OF INSTRUMENTAL PARAMETERS ON ESI-MS ANALYTE RESPONSE USING FACTORIAL DESIGN

6.1. Introduction

Variation of analyte response in electrospray ionization mass spectrometry (ESI-MS) constitutes a major concern in the quantitative applications of the technique (e.g. for metabolomics and interactomics). Despite several research efforts in this direction, a comprehensive model capable of predicting analyte solution concentration based on ESI-MS ion abundance remains elusive. Recent works have shown that analyte ESI-MS response may be modeled considering interrelationships of analyte physicochemical properties [153-156]. In order to better understand the way and manner in which analyte signal varies during the electrospray process, it is also important to understand which instrumental factors have the greatest effects on analyte response factors.

Several instrumental factors are known to influence analyte response in ESI-MS. In performing an ESI-MS experiment, it is important to identify the key parameters that elicit greatest influence on analyte signal intensity and quality. Although most instrument manufacturers often provide factory-built and tested software packages along with their instruments to assist the user in optimizing the instrument conditions during an experiment, it is still necessary for the user to be familiar with the way and manner in which the different parameters affect the results of a given experiment. More importantly, information on interactions between two or more parameters are often undetectable when using the factory installed optimization software; especially when one or more of the parameters involved in the interaction is a categorical variable, or cannot be varied automatically during the optimization procedure (e.g. the temperature of the capillary between the source and the mass analyzer). As such, a method that facilitates the identification of the main

effects due to individual instrument parameters as well as any interaction between them is useful for guiding optimization analysis.

Chemometrics-based techniques have been employed for optimizing experimental conditions in different fields of science for decades [161]. The most notable applications are centered on effective experimental designs aimed at screening for important factors, creating mathematical models for experiments, or optimizing selected factors that affect the results of an experiment. The prime advantage of using chemometrics-based experimental design techniques is the ability to gather useful information about the system by conducting a minimal number of experiments. The savings in time and resources realized by using these techniques is a strict consideration in most laboratories, and in industry, in particular.

Commonly employed chemometrics techniques include factorial designs (full factorial designs or fractional factorial designs), response surface designs, central composite designs, and Plackett-Burman designs, among others [161, 162]. In this part of the dissertation, full factorial experimental design technique has been employed to investigate the effects of four different instrument parameters on the ESI-MS analyte signal intensity of three GXG (Gly-Xxx-Gly; where Xxx represents the amino acids Phe, Glu, and Arg) tripeptides. Effects of spray voltage, capillary voltage, capillary temperature, and Q-array dc voltage (or tube lens voltage) are herein investigated. The experiments were carried out on two different ESI-MS instruments: LCQ Deca XP (Thermo Electron Corporation) and LCMS 2010 (Shimadzu). Yates' method was used to estimate the effect of each parameter as well as combinations of parameters. The most important parameters were identified by plotting the effects against their cumulative probability distribution on a normal plot. Statistical significance of each effect was also evaluated by analysis of variance (ANOVA).

6.2 Results and Discussions

The ESI-MS response factor of an analyte can be regarded as made up of two components: analyte-dependent factors and instrumental factors. Considerable amount of research studies have been carried out on evaluating the effects of analyte properties on ESI-MS response [157-159]; but many of the results in this area are still quite subjective. The effects of instrument parameters on analyte response have also been studied to a fair degree of details. However, most of these works are focused on individual parameters, or when multiple parameters are studied, their effects on analyte signals are interpreted on a case by case basis, with little or no mention made of likely (or sometimes *de facto*) interaction between parameters. The objective of this study is to develop a quick and easy method of evaluating the influence of instrumental parameters on analyte response during ESI-MS experiments, as well as identify occurrence and extent of interaction between the studied instrumental parameters during ESI-MS experiments.

The responses (in standard order) for the LCQ Deca XP are shown in Table 6.1 and those for the LCMS 2010 are shown in Table 6.2 for all three analytes at 50 μM each. An example of the final table obtained from Yates' algorithm is presented in Table 6.3. The numerical values of the effect of each treatment are presented in the seventh column in Table 6.3. These are the estimated effects calculated using Yates algorithm [141]. Negative values indicate that increasing the value of this term will decrease the analyte response, and positive values indicate increase in analyte response upon increasing the corresponding term. The greater the absolute value of a term, the greater its influence on analyte response. Since all the values are far greater than zero, visual inspection will seem to suggest that they all have influence on analyte response. However, several methods can be used to show that not all the factors or treatments are important in determining the analyte response. One method is to use the sum of squares to compute ANOVA and use the variance ratios to determine the significance of each effect. An easier and quicker method is the normal probability plot method, which graphically shows the

distribution of the effects. Here, the cumulative probabilities of the effects are plotted against their numerical values on a normal probability graph. The underlying principle is based from the well known concept of normal distribution. If the effects are due to chance and random process, then they should lie very close to the center of the plot, and on a straight line. Any effect that is not caused by random error should then be distinctly set apart from the cluster and be positioned away from the center.

Table 6.1 Analyte responses from Thermo LCQ Deca XP (values arranged in Yates order)

Treatment #	GDG	GFG	GRG
1	1680596	8732094	3276638
3	974415	8628939	2782437
2	2185035	10570093	3962149
6	1705129	11903243	3837871
4	1585208	7651542	5228588
7	1114735	6166111	4879344
9	204929	943993	1169601
11	150653	745194	1189226
5	1314400	6874246	2533553
8	1140971	8529214	2802865
10	1869415	9067236	3352340
13	1654943	11160180	3780775
15	1077664	4607648	4167216
12	716390	3866278	3739017
14	277020	1182409	1716992
16	201551	889617	1697960

Table 6.2 Analyte responses from Shimadzu LCMS-2010 (values arranged in Yates order)

Treatment #	GDG	GFG	GRG
1	130859	432351	272009
3	276981	1037874	539197
2	5261	15244	37047
6	8743	33174	71721
4	77932	247921	280562
7	166447	600291	595746
9	2778	2665	7893
11	2716	3107	18761
5	120553	387147	262183
8	265503	984912	536581
10	9210	29681	78780
13	18265	75630	162014
15	58619	198332	247902
12	152081	536371	573505
14	3006	3007	23833
16	3408	3913	60356

Table 6.3 Results after applying Yates' algorithm to the responses for GDG

GDG	1	2	3	4	Divisor	Estimate	SS
65880	212520	222251	361924	701463	16	43841	3.08E+10
146641	9731	139673	339539	258673	8	32334	4.18E+09
3809	134484	220692	130121	-627417	8	-78427	2.46E+10
5922	5189	118847	128552	-243955	8	-30494	3.72E+09
43632	204399	82874	-332084	-184423	8	-23053	2.13E+09
90852	16293	47247	-295333	-77314	8	-9664	3.74E+08
2581	113038	85119	-125841	154373	8	19297	1.49E+09
2608	5810	43433	-118114	63883	8	7985	2.55E+08
62102	80761	-202790	-82578	-22385	8	-2798	3.13E+07
142297	2113	-129294	-101845	-1569	8	-196	1.54E+05
5685	47220	-188105	-35627	36751	8	4594	8.44E+07
10609	27	-107228	-41687	7727	8	966	3.73E+06
34950	80195	-78648	73495	-19267	8	-2408	2.32E+07
78088	4924	-47194	80877	-6059	8	-757	2.29E+06
2757	43138	-75271	31454	7382	8	923	3.41E+06
3052	295	-42843	32429	975	8	122	5.94E+04

The cumulative probabilities are calculated for each effect by first rearranging the effects in ascending order. The probability (P_r) of any effect is then obtained using the formula:

$$P_r = \frac{100(r - 0.5)}{N - 1} \qquad \text{(eq 1)}$$

where r is the rank of the effect after rearrangement in ascending order and N is the total number of experiments. The ranks and probabilities used for all the effects in this study are shown in Table 6.4. Once the effects are ranked, the normal plots can be generated by plotting the probabilities on the vertical axis and the numerical values of the effects on the horizontal axis. Figures 6.1 - 6.3 illustrate the application of this method.

Table 6.4 Cumulative probabilities of the effects

Rank (r)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Probability (%)	3.33	10.00	16.67	23.33	30.00	36.67	43.33	50.00	56.67	63.33	70.00	76.67	83.33	90.00	96.67

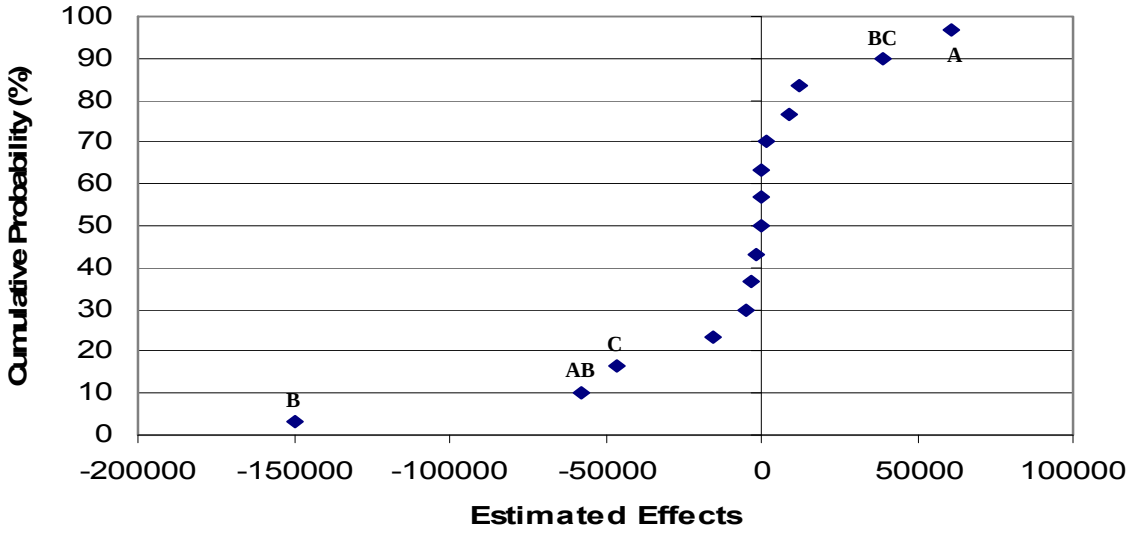

Figure 6.1: Normal Plot of cumulative probability against effects for GDG (LCMS-2010).

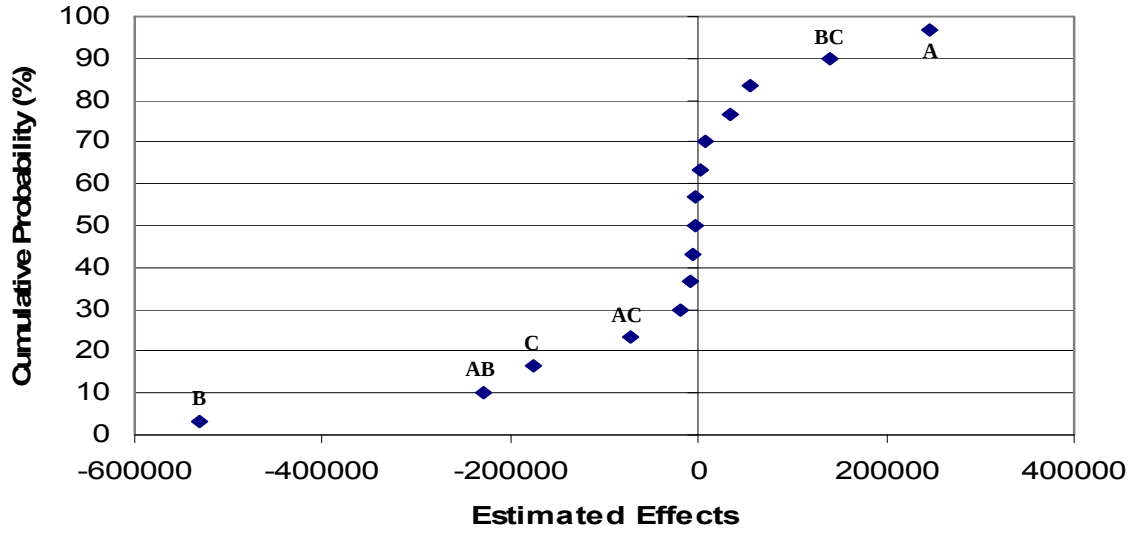

Figure 6.2: Normal Plot of cumulative probability against effects for GFG (LCMS-2010).

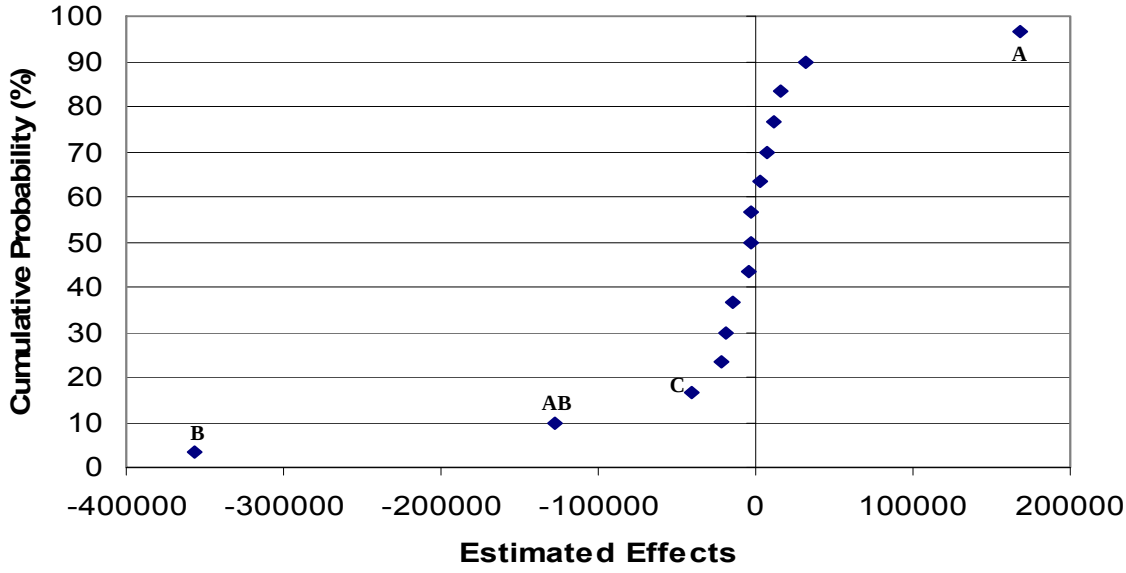

Figure 6.3: Normal Plot of cumulative probability against effects for GRG (LCMS-2010).

The main effects identified in these experiments include: A (spray voltage), B (tube lens or Q-array dc voltage), C (capillary temperature), and D (capillary voltage). Interactions between parameters can be two-termed, three-termed, or four-termed. The six two-term interaction effects are: AB (spray voltage with tube lens/ Q-array dc voltage); AC (spray voltage with capillary temperature); AD (spray voltage with capillary voltage); BC (tube lens or Q-array dc voltage with capillary temperature); BD (tube lens or Q-array dc voltage with capillary voltage); and CD (capillary temperature with capillary voltage). The four three-term interaction effects are: ABC (spray voltage, tube lens or Q-array dc voltage, and capillary temperature); ABD (spray voltage, tube lens or Q-array dc voltage, and capillary voltage); ACD (spray voltage, capillary temperature, and capillary voltage); and BCD (tube lens or Q-array dc voltage, capillary temperature, and capillary voltage). The only one four-term interaction effect possible is ABCD, incorporating all of the tested parameters.

The points labeled on the plots with their respective alphabetical notations clearly lie outside the straight line portion of the plots. For the LCMS 2010, the most important parameters, according to this method, are the spray voltage (A) and Q-array dc voltage (B). There appears to

be some level of interaction between both parameters (AB). For both GDG and GFG, the plots indicate that both capillary temperature (C), and an interaction term between capillary temperature and Q-array dc, are additional terms that may influence analyte response as well. It is interesting to note that these two terms (capillary temperature (C), and its interaction with Q-array dc) lie very close to the center on the normal plot for GRG, indicating that they are not as important as the other three effects earlier mentioned above. The reason for this slight discrepancy is not very clear. A plausible explanation may be that the analyte property is interacting with the instrumental factor. In other words, the effects of instrumental factors (often taken as constant for all analytes during ESI-MS) may in fact vary from analyte to analyte. This was indeed one of the features we set out to investigate from the inception of the experiment.

As a corollary, we deliberately chose the three analytes to be of different characteristics (acidic, neutral, and basic). At the pH of analysis (slightly below 6), GDG exists chiefly as a negatively charged species, GFG as a zwitterion, and GRG as a positively charged species. The percent distributions of species (calculated using ACD/Labs pK_a Predictor Ver.9, Toronto, ON) for the three analytes at different pH values are as shown in Figures 6.4a – 6.4c. The positive charge on the GRG at this pH is expected to be on the guanidinium group (on the arginine side chain) which has been reported to have a slightly hydrophobic property. This may contribute to its overall responsiveness during the ESI process. In other words, a combination of the essentially permanent positive charge on GRG and the potential hydrophobicity of the guanidinium group may favor the ionization process for GRG over its two counterparts which do not have these properties.

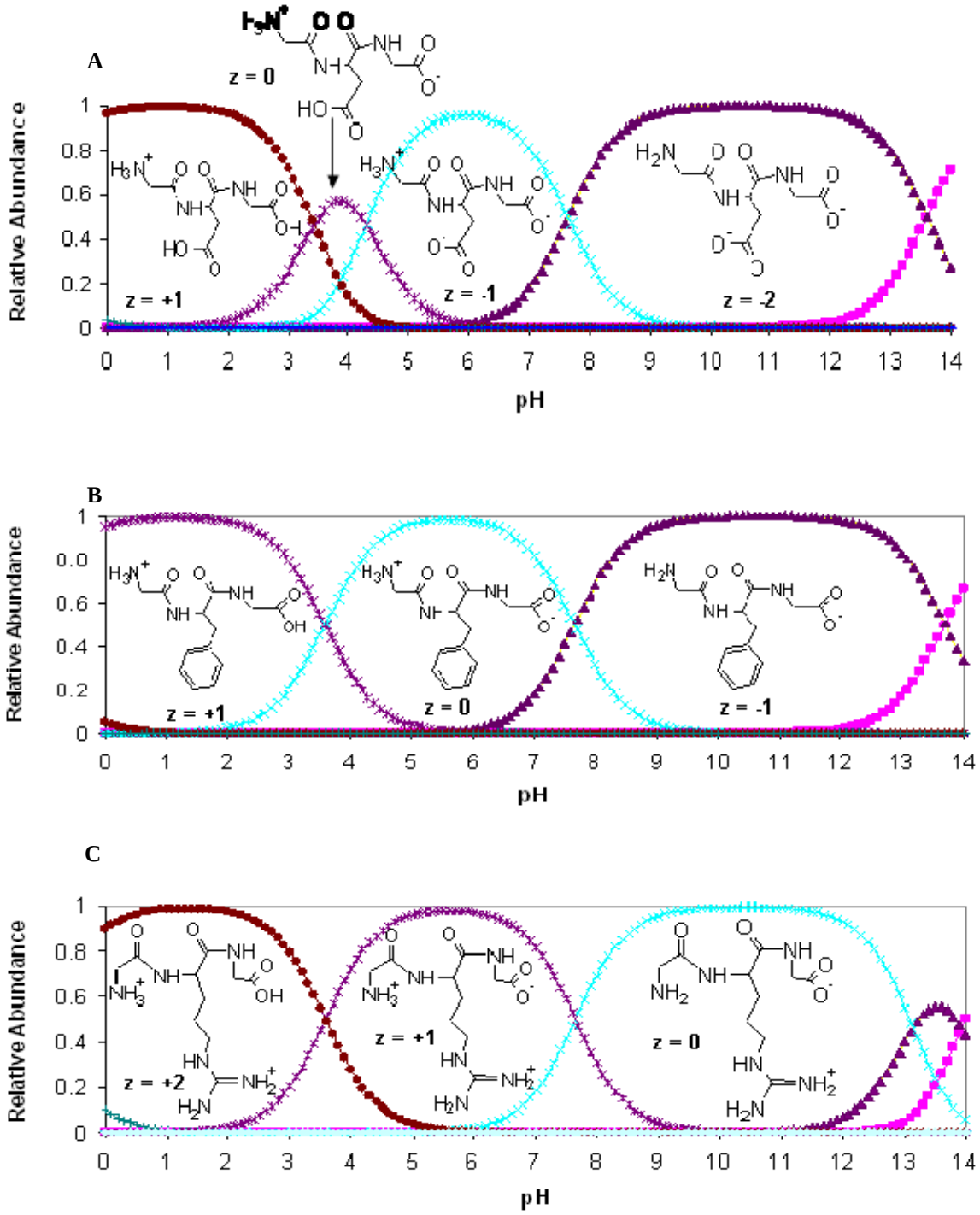

Figure 6.4: Charge state distributions vs pH for (a) GDG, (b) GFG, and (c) GRG

For the LCQ Deca XP data, the most important parameters are capillary temperature (C), tube lens voltage (B), and an interaction term (BC) between both capillary temperature and tube lens voltage. This is interesting in two senses. First is the fact that the instrument manufacturer actually indicated in the instrument manual that changing the capillary temperature during an experiment may require also changing the tube lens voltage in order to obtain optimum signal intensity; another way of saying that both parameters interact. Secondly, the plots indicate that spray voltage does not have as much effect on analyte response as it does in the case of the LCMS 2010. This does not by any means underestimate the importance of spray voltage during analysis. Rather, what it means is that once the voltage required for the onset of droplet formation is reached, any further variation in spray voltage will have little effect on ionization efficiency, and consequently analyte response. However, knowing fully well that too low a spray voltage may lead to inadequate droplet charging, and consequently sensitivity, while too high values may cause arcing, it seems more pertinent in view of the foregoing results to state that, within the boundary values of the factor space studied for this parameter on this instrument, there is no effect on analyte response.

Another noteworthy observation is the fact that, similar to the case of LCMS 2010, the same factors influence analyte response for GDG and GFG while GRG differs in one of the factors. All three analytes have BC (interaction of tube lens voltage with capillary temperature) in common; both GDG and GFG are also affected by capillary temperature while GRG is affected by tube lens voltage (similar to what was observed in the LCMS 2010). This observation lends further credibility to the above stated assumption that the analyte property seems to interact with the capillary temperature during ESI-MS analyses.

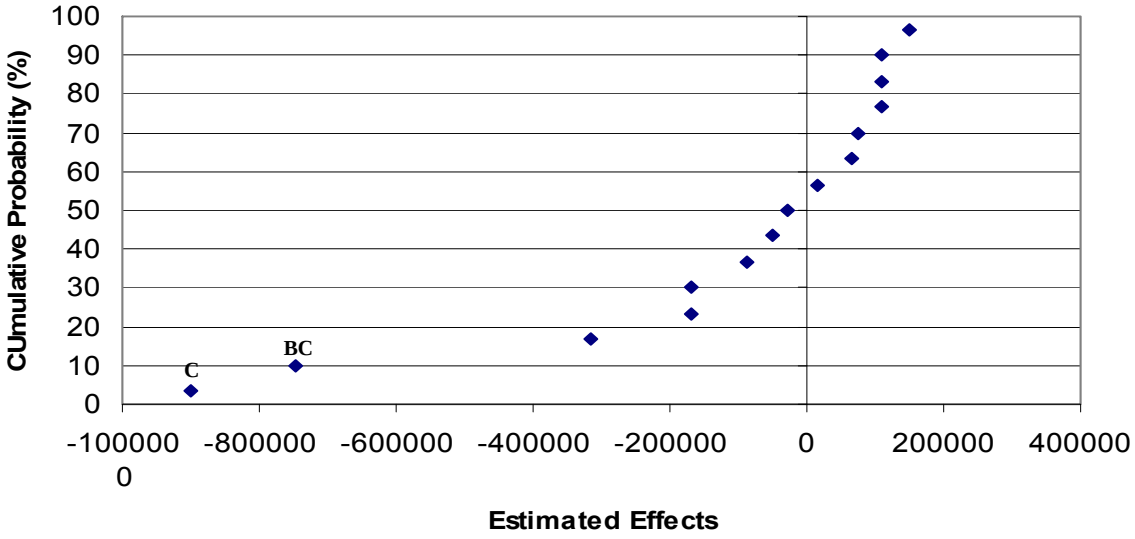

Figure 6.5: Normal plot of cumulative probability against effects for GDG (LCQ Deca XP)

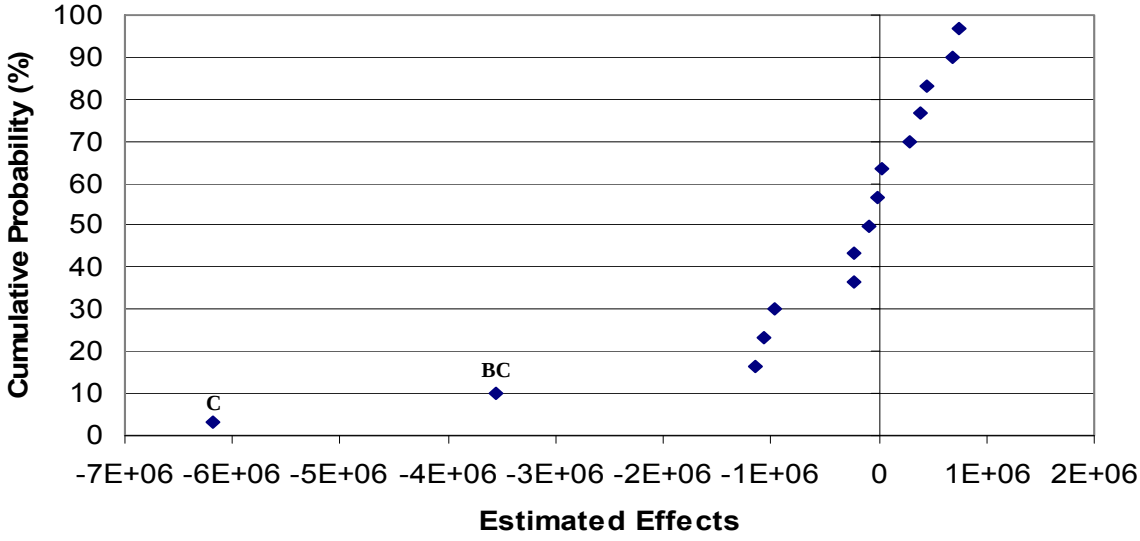

Figure 6.6: Normal plot of cumulative probability against effects for GFG (LCQ Deca XP)

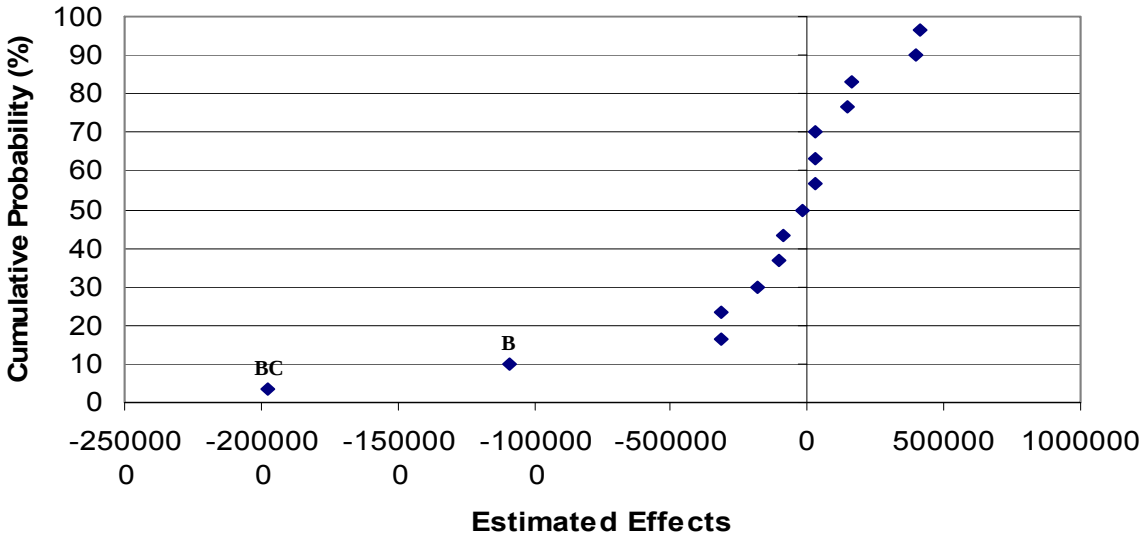

Figure 6.7: Normal plot of cumulative probability against effects for GRG (LCQ Deca XP)

The results of the three replicate measurements made at 25 μM on the LCMS 2010 instrument are exactly the same as those obtained at 50 μM. The absolute numerical vales of the effects at 25 μM are about half of those at 50 μM but the order of the most important parameters is exactly the same in both cases. Very slight differences were observed in the ordering of the factors that lie on the straight line in the center of the normal plots between the 25 μM and 50 μM measurements. However, since these effects are not significant, such differences do not matter; and it can be concluded that the method is not subject to concentrations effects.

Particular attention needs to be paid to the interaction effects (AB for the LCMS 2010 and BC for the LCQ Deca XP). For the LCQ Deca XP, the instrument manual gives information on how both parameters may behave with each other during an ESI-MS experiment. However, such information on factor interaction is not available on the LCMS 2010. The extent of interaction between the spray voltage and the Q-array dc voltage is depicted graphically in Figure 6.8.

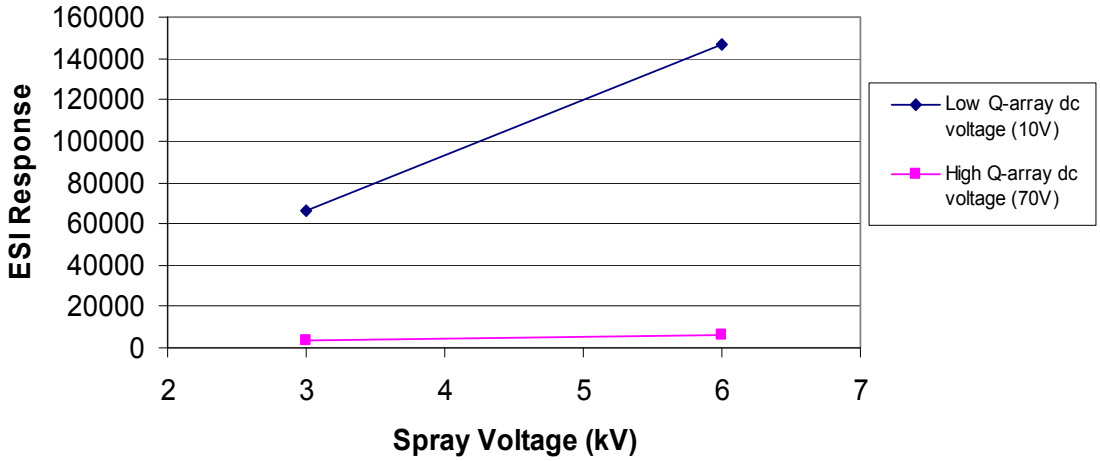

Figure 6.8: Interaction between spray voltage and Q-array dc voltage in LCMS-2010 (GDG)

If there is no interaction, the lines should be parallel. The divergence between the two lines clearly indicates that there is indeed considerable interaction. The same type of plot for the LCQ Deca XP (Fig. 6.9) reveals little or no interaction between these two parameters. It is clear that the two lines will not diverge significantly going towards the factory limit of 8 kV for this factor.

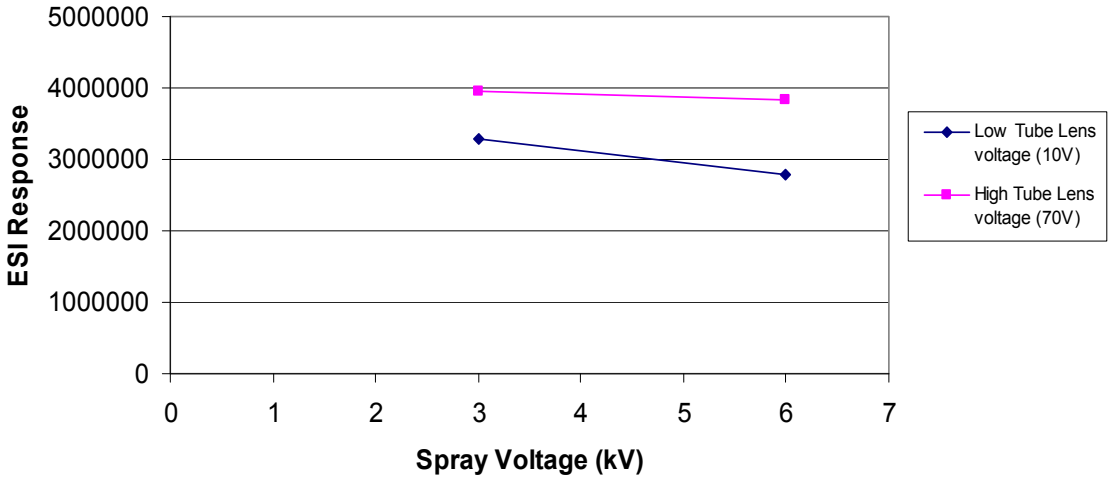

Figure 6.9: Interaction between spray voltage and tube lens voltage in LCQ Deca XP (GRG)

6.3 Analyses of Residuals

In order to assess the statistical significance of the effects of the selected parameters, particularly to further ascertain that the interaction effects observed are important, the numerical values of the effects were used to compute a predicted response for each of the runs based on model equations (e.g. for GDG on the LCMS 2010 in Equation 2 and for GRG on the LCMS 2010 in Equation 3):

$$Y = 43841.5 + \left(-78427.2/2\right)x_B + \left(-30494.4/2\right)x_{AB} + \left(-23052.8/2\right)x_C + \left(19296.58/2\right)x_{BC} + \left(32334.17/2\right)x_A \quad \text{(eq 2)}$$

$$Y = 107355.7 + \left(-161314/2\right)x_B + \left(-58740.8/2\right)x_{AB} + \left(78724.83/2\right)x_A \quad \text{(eq 3)}$$

In these equations, Y represents the predicted response, while x_A, x_B, x_C, x_{AB}, and x_{BC} represent the contrast of the effects in the design matrix. The first term in both equations is the mean effect. Each effect is divided by two to account for the high and low levels used for each factor. The predicted responses calculated from these model equations, alongside their residuals, are presented in Tables 6.5 and 6.6 for GDG and GRG, respectively. A normal plot of the residuals is also presented in Figure 6.10 for GDG using model equation 2. The calculation of predicted value was repeated with terms C and BC excluded in order to verify the significance of these effects. The normal plot of the resulting residuals is shown in Figure 6.11. It is clear that the fit is better when the two terms are included in the model than when they are excluded. This indicates that the capillary temperature and its interaction with the Q-array dc are important in determining the response of both GDG and GFG. In the case of GRG, the model in equation 3 gave better fit than when the capillary temperature and any of its interactions were included in the model. The normal plots of the residuals for GRG are presented in Figures 6.12 and 6.13.

Table 6.5 Predicted responses from model equation and the residuals for GDG on LCMS 2010

Treatment #	Observed Response	Predicted Response	Residuals	% of Obs. Res
1	65880	72816	-6936	10.5
3	146641	135644	10997	7.5
2	3809	5586	-1777	46.7
6	5922	7426	-1504	25.4
4	43632	30466	13166	30.2
7	90852	93295	-2443	2.7
9	2581	1830	751	29.1
11	2608	3670	-1062	40.7
5	62102	72816	-10714	17.3
8	142297	135644	6653	4.7
10	5685	5586	99	1.7
13	10609	7426	3183	30.0
15	34950	30466	4484	12.8
12	78088	93295	-15207	19.5
14	2757	1830	927	33.6
16	3052	3670	-617	20.2

Table 6.6 Predicted responses from model equation and the residuals for GRG on LCMS 2010

Treatment #	Observed Response	Predicted Response	Residuals	% of Obs. Res
1	124587	119280	5307	4.3
3	283682	256746	26936	9.5
2	16755	16707	48	0.3
6	35291	36691	-1400	4.0
4	122251	119280	2971	2.4
7	242020	256746	-14726	6.1
9	4425	16707	-12281	277.5
11	8185	36691	-28506	348.3
5	120264	119280	984	0.8
8	277553	256746	20808	7.5
10	33397	16707	16690	50.0
13	78235	36691	41544	53.1
15	110017	119280	-9263	8.4
12	223726	256746	-33019	14.8
14	12250	16707	-4456	36.4
16	25053	36691	-11638	46.5

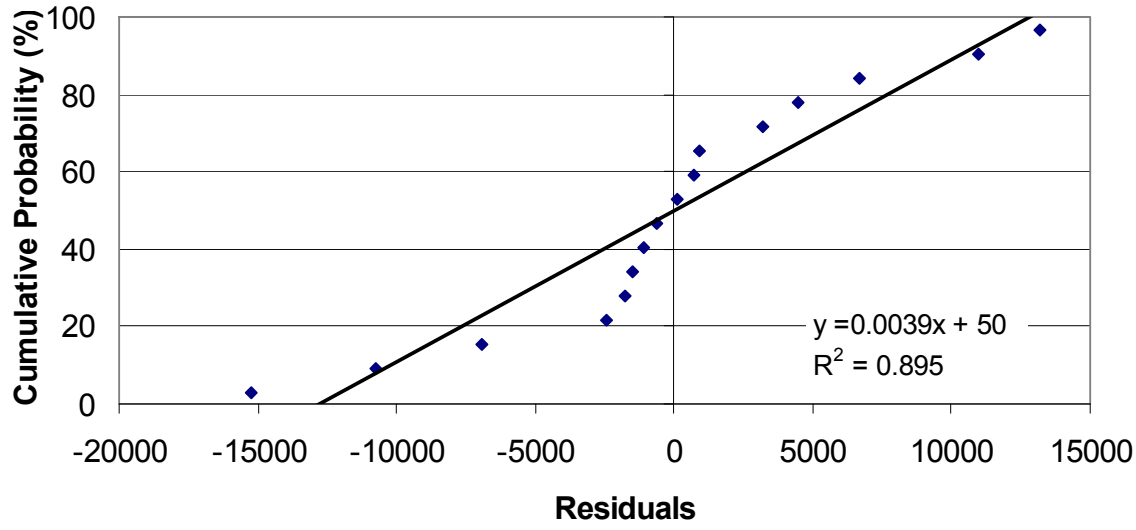

Figure 6.10: Normal plot of residuals for GDG obtained from model equation on LCMS 2010

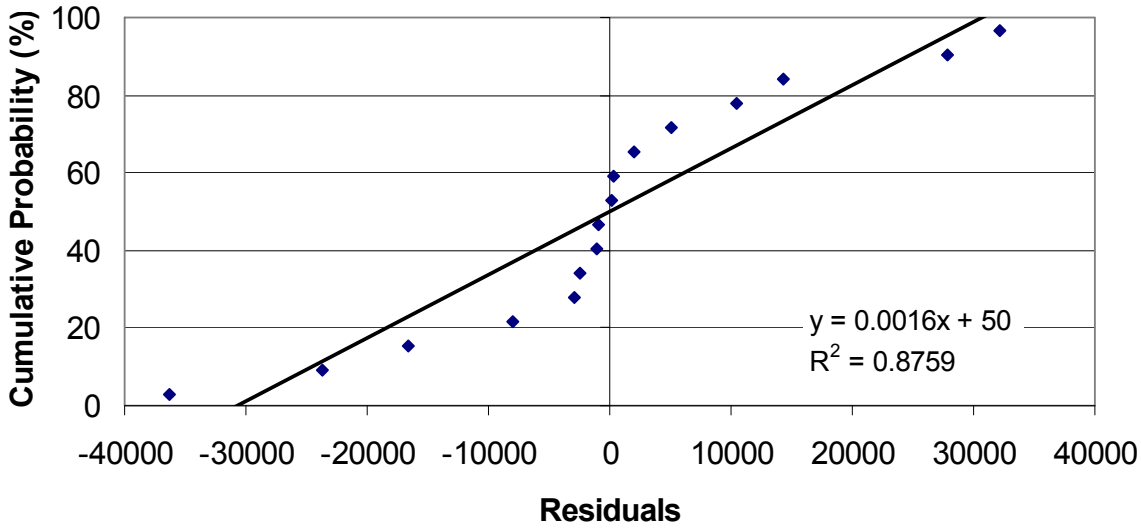

Figure 6.11: Normal plot of residuals for GDG (excluding the effects of capillary temperature and its interactions) on LCMS 2010

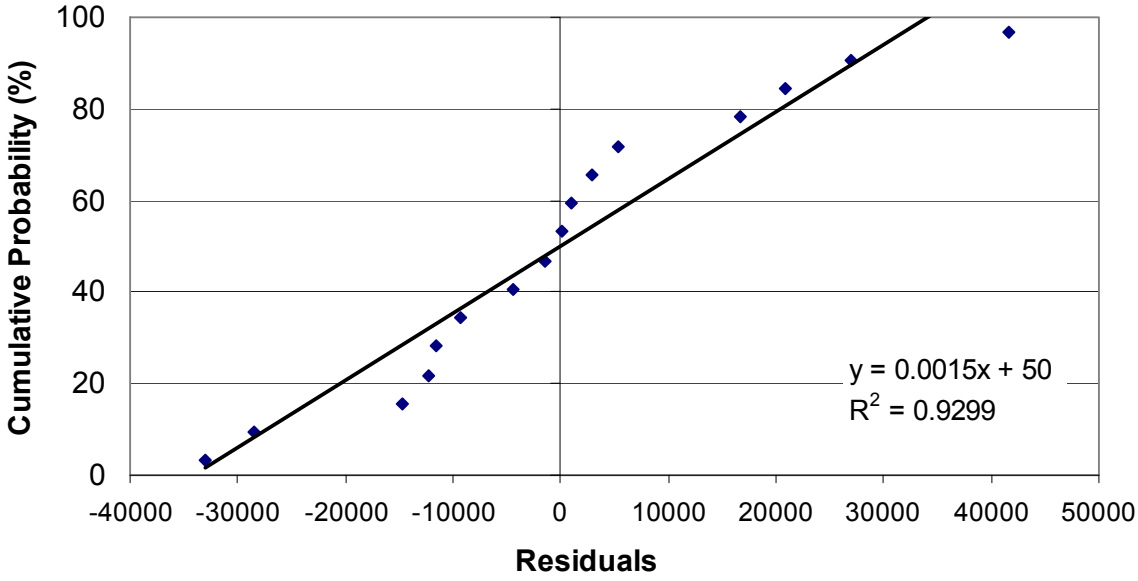

Figure 6.12: Normal plot of residuals for GRG obtained from model equation on LCMS 2010

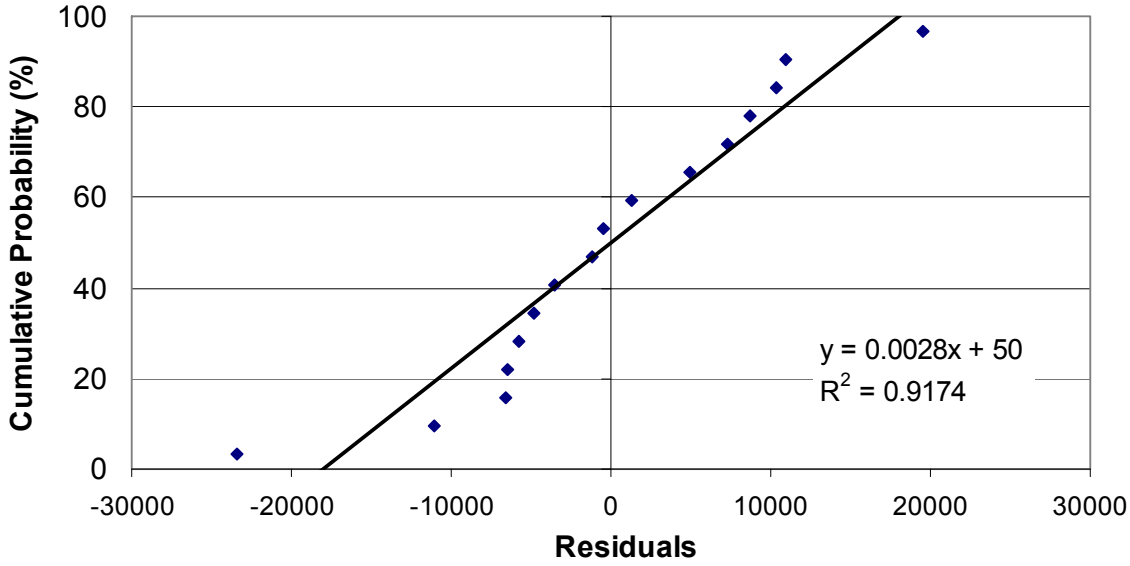

Figure 6.13: Normal plot of residuals for GRG (including the effects of capillary temperature and its interactions) on LCMS 2010)

A comparison of the correlation coefficient of the normal plots of residuals indicate that the model in Equation 2 predicts the response better for GDG than a model that excludes the effects of capillary temperature and its interaction with Q-array dc voltage on the LCMS 2010 instrument. The same result was obtained for GFG (not shown). These further support the results of the normal plots of effects, where it was shown graphically that these two terms do not lie on the straight line in the center with the rest of the unimportant terms (Figures 6.1 and 6.2 above, and the interaction plot in Figure 6.8). For GRG, Eq 3 predicts the response better than any other model that includes any other term, including the capillary temperature and any of its interactions. The percent error in both models is less than 20% for more than half of the runs.

In an attempt to further justify the statistical significance of each of these effects, an ANOVA was carried out on the sum of squares (last column in Table 6.3). In order to construct ANOVA, an estimate of the residual error sum of squares is required. One way to obtain this is to take an average of the squares of residuals (difference between observed responses and predicted responses). Another rough estimate is the average of the sum of squares of the effects that lie mostly on the straight line on the normal plot. The residual sum of squares can also be calculated as the difference between the total sum of squares (a summation of all the squares of replicate measurements of responses) and the treatment sum of squares. All three methods were used to calculate residual error sum of squares. There are slight differences in the results. Since not all the experiments were carried out in replicates, we have used the first two methods to estimate the ANOVA in this study. The variance ratios obtained for GDG and GRG are presented in Tables 6.7 and 6.8. Variance ratio 1 is the ratio of the treatment sum of squares to the mean residual sum of squares obtained by pooling the sum of squares for the unimportant effects and averaging them (R1). Variance ratio 2 is calculated in a similar way but using the mean residual sum of squares calculated as the average of the squares of residuals (*vide supra*). F-test is then used to compare the variance ratios to the tabulated value at P = 0.05 level and the respective degrees of freedom. The degree of freedom for each treatment sum of squares is 1. The degree

of freedom for the residual error sum of squares varies depending on the way the residual errors are calculated. If the first method is used, where the sums of squares of all unimportant effects are averaged, then the degrees of freedom will be the number of these effects that are used. If on the other hand the average of the squares of the residuals is used, then the number of degrees of freedom is N-1 (where N is the number of treatments). The tabulated F value at this probability level, and at (1, 8) degrees of freedom (for variance ratio 1) and (1, 15) degrees of freedom (for variance ratio 2), are 5.317 and 4.543 respectively for GDG. Tabulated F values for GRG for both methods are 4.844 and 4.543, respectively.

Table 6.7 Variance ratios of the treatment sum of squares to residual sum of squares (GDG)

Treatment Number	Treatment Sum of Squares	Variance Ratio 1 (Treatment SS/R1)	Variance Ratio 2 (Treatment SS/R2)	Effects
1	30753175501	660.787	633.083	**1**
3	4181993336	89.858	86.090	**A**
2	24603281885	528.646	506.482	**B**
6	3719637791	79.923	76.572	**AB**
4	2125732499	45.675	43.760	**C**
7	373590912	8.027	7.691	**AC**
9	1489432513	32.003	30.661	**BC**
11	255062194	5.480	5.251	**ABC**
5	31317081	0.673	0.645	**D**
8	153925	0.003	0.003	**AD**
10	84413219	1.814	1.738	**BD**
13	3731980	0.080	0.077	**ABD**
15	23200278	0.498	0.478	**CD**
12	2294720	0.049	0.047	**ACD**
14	3405870	0.073	0.070	**BCD**
16	59373	0.001	0.001	**ABCD**

R1 = 46540195 (8 degrees of freedom); R2 = 48576847 (15 degrees of freedom).

Table 6.8 Variance ratios of the treatment sum of squares to residual sum of squares (GRG)

Treatment Number	Treatment Sum of Squares	Variance Ratio 1 (Treatment SS/R1)	Variance Ratio 2 (Treatment SS/R2)	Effects
1	184403826647	768.676	516.306	**1**
3	24790397533	103.337	69.410	**A**
2	104088503756	433.887	291.434	**B**
6	13801902842	57.532	38.643	**AB**
4	3075682195	12.821	8.611	**C**
7	1051694090	4.384	2.945	**AC**
9	2025878	0.008	0.006	**BC**
11	81420544	0.339	0.228	**ABC**
5	117180625	0.488	0.328	**D**
8	47196900	0.197	0.132	**AD**
10	980597910	4.088	2.746	**BD**
13	116694006	0.486	0.327	**ABD**
15	188847145	0.787	0.529	**CD**
12	28928262	0.121	0.081	**ACD**
14	13722085	0.057	0.038	**BCD**
16	10569001	0.044	0.030	**ABCD**

R1 = 239897859 (11 degrees of freedom); R2 = 357159915 (15 degrees of freedom).

Those effects with variance ratios greater than the tabulated values are the significant ones. It is clear that most of the effects identified as important by the normal plots also have their variances ratios greater than the tabulated values from F-test. For GDG, the statistically significant effects and their order are B>A> AB> C> BC>AC (the effects have been described earlier). For GRG, the important effects are B>A>AB>C. Although C (capillary temperature) was not included in the model (Eq 3) used to calculate the predicted responses for GRG, and hence the residuals, the above results show that this parameter does have some effect on analyte response (its variance ratio is almost twice the tabulated value), albeit very small compared to the other selected parameters. A close inspection of the normal plot in Figure 6.3 shows that this point is slightly off the center.

6.4 Five-variable model incorporating the rf component of Q-array voltage in LCMS 2010

The Q-array voltage has two voltage components: dc voltage and rf voltage. It is conceivable that both voltages may interact with each other and with other parameters in an ESI-MS experiment. A five-variable model was therefore designed to investigate the possibility of these interactions. It was found that there are several interaction terms resulting from the inclusion of the fifth parameter into the model as shown in the normal plots in Figures 6.14-6.16 below.

For all three analytes, Q-array dc voltage, Q-array rf voltage, spray voltage, and the two-term interactions between these parameters seem to dominate the effects on analyte ESI response. A three term interaction between all three parameters also seems to be very important as it ranks higher than both capillary temperature and capillary voltage. It is observed that in all the experiments performed in both 4-variable and 5-variable models, capillary voltage (CDL voltage in LCMS 2010) did not have much influence on analyte ESI response.

Another observation in the 5-variable model is that the capillary temperature, as well as its interaction terms, lies closer to the center of the normal plots for GRG than for GFG and GDG.

This is similar to the scenario in the 4-variable model. This trend further corroborates the conclusion that capillary temperature has less effect on the ESI response of GRG (a basic peptide) than it has on GDG and GFG, which are acidic and neutral, respectively. A possible explanation for this observation was given earlier and linked to the reported hydrophobicity of the guanidinium group on the arginine residue of GRG.

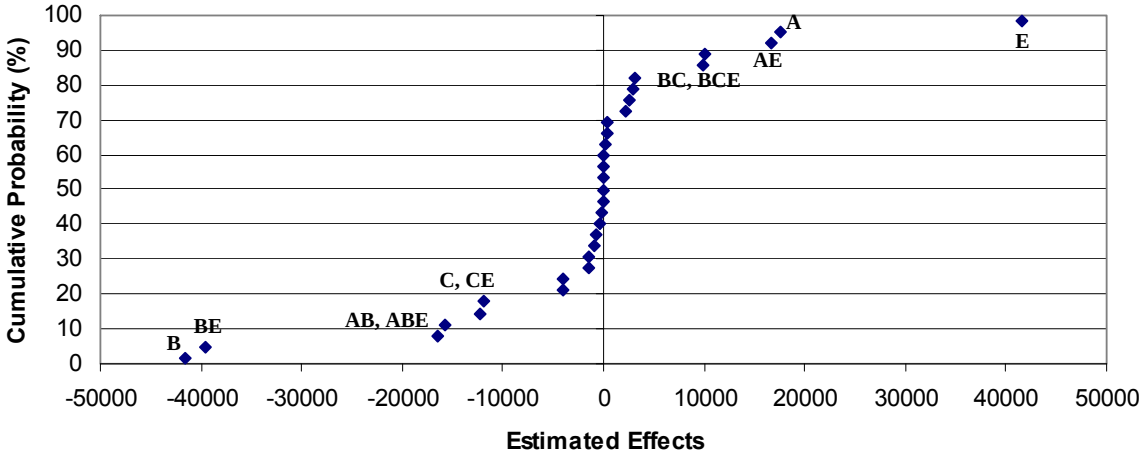

Figure 6.14: Normal plots of effects for GDG for 5 variables

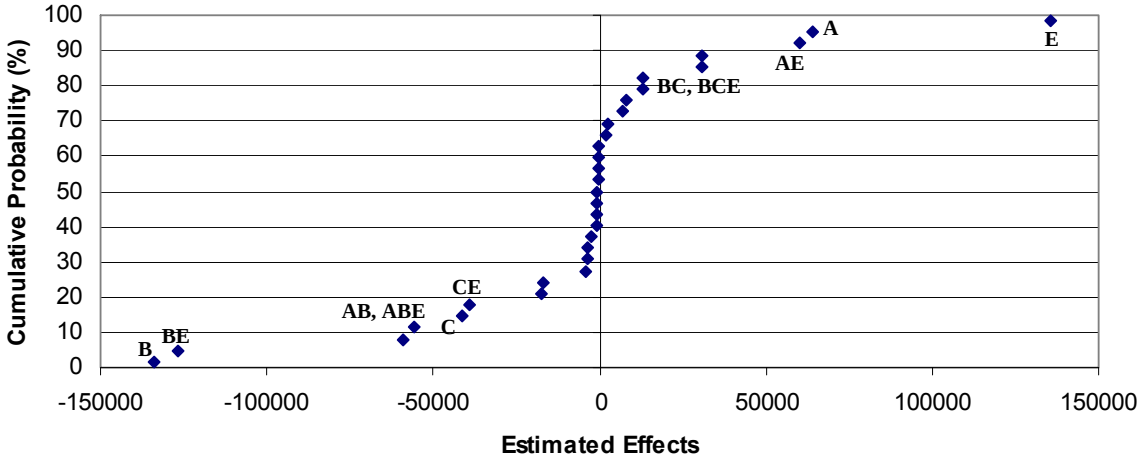

Figure 6.15: Normal plots of effects for GFG for 5 variables

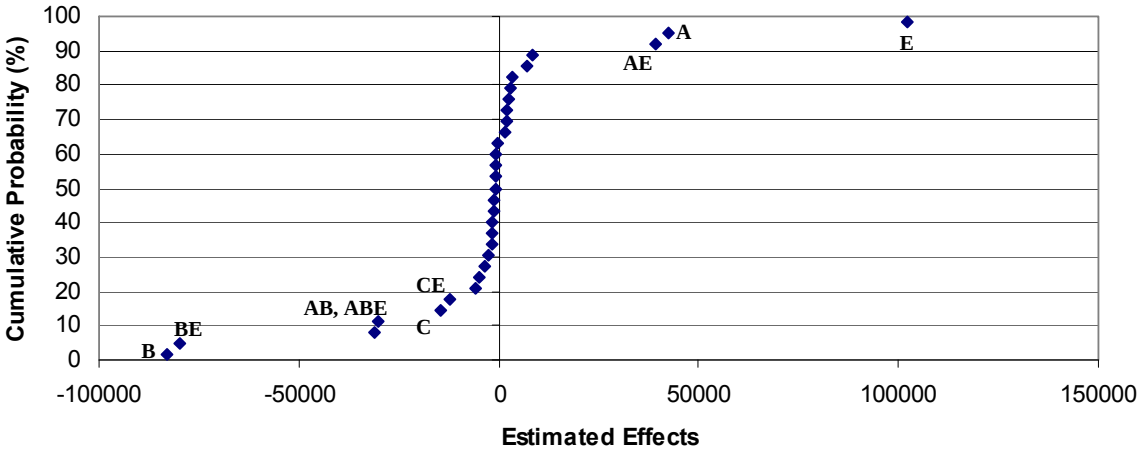

Figure 6.16: Normal plots of effects for GRG for 5 variables

CHAPTER 7

CONCLUSIONS

"Life is not about answers; it is about questions and the quest to find solutions to stated problems"

- Richard N Zare

The work presented in this dissertation comes as a package that details the application of electrospray mass spectrometry for studying noncovalent interactions in a biological system that has hitherto not been studied using this technique, as well as provides different methods that can be used to investigate the effects of both instrument and analyte parameters on analyte response in order to gain better understanding of the role of these parameters during the ESI process.

Results from the work on integrin fragments binding studies have demonstrated in it is possible to obtain binding information for complexes formed between integrin fragments and RGD peptides by ESI-MS techniques. To the best of our knowledge, this is the first application of ESI-MS for studying noncovalent interactions between different peptide fragments representing the integrin binding regions and RGD peptides.

Out of five integrin fragments studied with the RGD ligands, only the α_V and α_{IIb} fragments showed appreciable binding with some of the ligands. Amongst the thirteen RGD-based ligands studied, the highest binding affinities with the α_V and α_{IIb} fragments were observed for GRGDNP and GRGDSP. None of the cyclic peptides studied (including Cilengitide, the famed Kessler peptide) showed any appreciable level of affinity for the studied integrin fragments. Also, the fibronectin fragment (fibronectin is one of the extracellular matrix proteins that bind to integrins via the RGD motif) did not show any appreciable binding with the integrin fragments. The lack of binding between the integrin fragments and these peptides may be due to the absence (in the integrin fragments) of a requisite protein tertiary conformation for the binding process. As mentioned in the introduction section, certain key regions of both the α and the β

subunits of the whole integrin protein molecule are vital to the ligand recognition and binding process. Future study of RGD-ligand binding to whole integrin protein molecules or subunits will be possible with mass spectrometry instruments having mass analyzers that possess high resolution and high mass range. Nevertheless, the results of this work serve as proof of concept as well as help establish some fundamentals of ESI-MS application for the integrin-RGD system. This work also lays the foundation for future development of a rapid and high throughput screening method for synthetic peptidomimetic drug molecules that target the integrin protein molecules.

With respect to the workability of this approach in real biological systems, one of the major challenges envisaged is the marked difference in the behavior of molecules and ions in a physiological environment as compared to the ESI conditions employed in this study. To employ more physiologically relevant conditions in ESI-MS, though challenging, may lend greater merit to the use of ESI-MS for obtaining affinity values relevant to physiological conditions; as well as better correlation to measurements made by complementary solution phase techniques. This challenge constitutes a major future goal as it pertains to this work.

Another major challenge, which can be regarded as the bane of most ESI-MS-based binding affinity studies, is the issue of analyte response factor during the ESI-MS process. In any binding affinity study, equilibrium concentrations of participating species are needed to calculate binding constants. This information is not readily available from the ESI-MS experiments in host-guest complexation. For the host and the guest species, a calibration plot can be obtained from measurements of ion abundances at different concentrations of each species measure individually. A response factor, taken as the slope of such calibration plots, can be defined as the correlation coefficient between ion abundance and equilibrium solution concentration of each species. Provided the response factor is time invariant, and is not affected by the presence of other species during the electrospray process, it can then be used to estimate equilibrium concentration of the participating species in a host-guest complexation experiment. This is

practical for free host and guest molecules, but not for the complex. In most cases, response factors for free host and host-guest complex are assumed to be similar. This assumption may only be valid under certain conditions (e.g. large host binding small guest, or equivalent charge states between free and complex ions). In many instances, these assumptions do not hold; and binding constants so determined by ESI-MS do not reflect the true solution binding affinities. The quest for a solution to this problem gave birth to the other experiments described in this dissertation pertaining to the use of different statistical methods for evaluating the dependence of analyte response factor on instrument parameters as well as analyte properties.

Understanding the physicochemical factors affecting ESI-MS response of any species is the first step towards developing an effective method for evaluating its response factor. The results of the work in this part of the dissertation have demonstrated the possibility of employing statistical tools for delineating the parameters that are most relevant to response factor determination. Using the p-values, it was possible to identify the analyte physicochemical parameters that are most pertinent to response factor determination at different pH conditions. Although little can be said about this phenomenon at the present time, the fact that qualitative treatment could not categorically pinpoint the parameters that have most influence on response factor determination at different pH conditions, whereas statistical treatment achieved this to a reasonable extent, is another reason to consider statistical treatment for producing an ESI-MS response factor model. Such is the case with Log D. Qualitative data treatment did not regard this parameter to be substantially important (despite reasonable expectation that it should have some correlation with response) whereas it was accorded significant relevance by both quantitative regression models employed. Another advantage of statistical methods is the ability to evaluate the effects of multiple parameters simultaneously, compared to other methods where single parameters have been individually evaluated for correlation with response factor. In addition, the ability of decision tree regression models to handle data sets that might exhibit nonlinear relationships compared to general multiple linear regression models is noteworthy. This is

highlighted by the large difference in the magnitude of SSE between MLR and regression tree analyses across the three different pH conditions tested.

The efficacy of a novel dynamic response factor determination technique developed in our lab was demonstrated. This technique is based on the flow injection band-broadening of an injected plug of sample in a long tubing before entering the ESI-MS ion source. The technique offers a number of advantages, among which is the ability to measure response factors of multiple analytes within a short period of time, using a single solution containing a small amount of sample. Compared to the usual method, where a calibration plot is generated by measuring responses at different discrete concentrations of the analyte, this is a great improvement in speed and material consumption. Additionally, several replicate measurements of the response factors can be made quickly, thereby reducing random temporal variation of response.

Although the results from this work have demonstrated the applicability of statistical methods for identifying pertinent physicochemical parameters that influence response factors, it should be noted that the quest for a comprehensive model that fully describes the influence of all possible physicochemical parameters on analyte ESI-MS response factor is far from over. The analyte set used in this study is far from being representative of the entire spectrum of analytes studied with ESI-MS. Future works in this direction will involve the extension of the models to other more encompassing sets of (small molecule) analytes, while also considering other statistical data treatment techniques.

Having explored the statistical evaluation of the effects of analyte physicochemical properties on ESI-MS response, the only questioned that remains to be answered is that of the dependence of analyte response on instrument parameters. A good understanding of the way and manner in which instrument parameters influence analyte response in ESI-MS is essential in order to obtain optimum response and reproducible data. Particular attention should be paid to how parameters are changed from run to run as possible interactions between parameters may lead to irreproducible results. The implication of inter-parameter interaction is that it becomes

statistically impracticable to interpret the effect of one parameter independently of the other. Although most mass spectrometry instruments come with tuning softwares that can be used to optimize certain parameters prior to acquiring data, these softwares may not account for interactions between parameters that can only be changed by the user. An example is the case of LCQ Deca XP, where a strong interaction was found between capillary temperature and tube lens voltage. The tuning software will optimize the tube lens voltage at the temperature of the capillary at the time of tuning. However, this value will change (and may become less than optimum) when the capillary temperature is manually varied following tuning.

The results of the chemometrics experiments described in this dissertation have demonstrated how simple factorial designs can be used to elucidate the influence of instrumental parameters on analyte ESI response. It was shown that this basic technique can be used to identify otherwise unnoticeable interactions between two or more parameters. Both graphical and statistical methods were used to justify selection of important parameters. The findings in this study are very important and need to be given considerations in the design and optimization of mass spectrometric experiments on these instruments. It is noteworthy to mention that this technique is limited to designs with few factors being investigated. The introduction of higher order interactions (e.g. three-term, four-term, etc.) as the number of factors increases renders the procedure cumbersome when used to study many parameters. Fractional factorial designs can be used to study multiple parameters where it is difficult to apply full factorial design. However, important information about interactions is often lost due to the removal of factors inherent to performing fractional factorial designs. Also, factorial designs are generally best for quick identification of important effects and not necessarily for optimizing the factors as they do not account for curvature in the levels of the factors. Other chemometrics tools, such as central composite (or response surface) designs, simplex centroid, and simplex lattice designs are suggested for optimization.

REFERENCES

1. Bohm, H-J.; Schneider, G. *Protein-Ligand Interactions*Wiley-VCH GmbH & Co. KGaA. Weinheim. 2003. pp1-7.
2. Ganem, B; Li, Y.T; Henion, J.D. *J. Am. Chem. Soc.* **1991**, *113*, 6294-6296.
3. Daniel, J.M.; Friess, S.D.; Rajagopalan, S.; Wendt, S.; Zenobi, R. *Int. J. Mass Spectrom*. **2002**, *216*, 1-27.
4. Brodbelt, J.S. Int. *J. Mass Spectrom.* **2000**, *200*, 57–69.
5. Schalley, C.A. *Int. J. Mass Spectrom.* **2000**, *194*, 11–39
6. Loo, J.A. *Mass Spectrom Rev.* **1997**, *16*, 1–23.
7. Fenn, J.B.; Mann, M.; Meng, C.K.; Wong, F.S.; Whitehouse, C.M. *Science*. **1989**, *246,* 64.
8. Fenn, J.B.; Mann, M.; Meng, C.K.; Wong, F.S.; Whitehouse, C.M. *Mass Spectrom. Rev.* **1990**, *9*, 37.
9. Plow F.E.; Haas, T.A.; Zhang, L.; Loftusi, J. Smith, J.W. *J. Biol. Chem.* **2000**, *275*, (29), 21785–21788.
10. Garret, R.H.; Grisham, C.M. *Biochemistry*. Third edition. Thomson Brooks/Cole. Belmont, CA, 2005, pp. 267.
11. Garret, R.H.; Grisham, C.M. *Biochemistry*. Third edition. Thomson Brooks/Cole. Belmont, CA, 2005, pp. 1077.
12. Albelda, S.M.; Buck, C.A. *The FASEB Journal*. **1990**, *4*, 2868-2880.
13. Haubner, R.; Finsinger, D.; Kessler, H. *Angew Chem.* Int Ed. Engl. **1997**, *36*, 1374-1389.
14. Fenn, J.B.; Thumshirn, G.; Hersel, U.; Goodman, S.L.; Kessler, H. *Chem. Eur. J.* **2003**, *9*, 2717-2725.
15. Ruoslahti. E. *Annu. Rev. Cell Dev. Biol.* **1996**, *12*, 697-715.

16. Kumar, C.C.; Armstrong, L.; Yin, Z.; Malkowski, M.; Maxwell, E.; Ling, H.; Yaremko, B.; Liu, M .; Varner, J.; Smith, E.M.; Neustadt, B.; Nechuta, T. *Adv. Exp. Med. Biol.* **2000**, *476,* 169.

17. Curley, G.P.; Blum, H.; Humphries, M.J.; *Cell Mol. Life Sci.* **1999**, *56*, 427.

18. Eliceiri, B.P.; Cheresh, D.A. *Cancer J.* **2000**, *13*, 245-249.

19. Friedlander, M.; Theesfeld, C.L.; Sugita, M.; Fruttinger, M.; Thomas, M.A.; Chang, S.; Cheresh, D.A. *Proc. Natl. Acad. Sci.* **1996**, *93,* 9764-9769.

20. Dechanstreiter, M.A.; Planker, E.; Matha, B.; Lohof, E.; Holzenmann, G.; Jonczyk, A.; Goodman, S.L.; Kessler, H. *J. Med Chem.* **1999**, *42*, 3033-3040.

21. Burke, P.A.; DeNardo, S.J.; Miers, L.A.; Lamborn, K.R.; Matzku, S.; DeNardo, G.L. *Cancer Res.* **2002**, *62,* 4263-4272.

22. Marinelli, L.; Gottschalk, K.-E.; Meyer, A.; Novellino, E.; Kessler, H. *J. Med Chem.* **2004**, *47*, 4166-4177.

23. Gurrath, M.; Muller, G.; Kessler, H. Aumailley, M.; Timpl, R. *Eur. J. Biochem.* **1992**, *210*, 911-921.

24. Marinelli, L.; Gottschalk, K.-E.; Meyer, A.; Novellino, E.; Kessler, H. *J. Med Chem.* **2004**, *47*, 4166-4177.

25. Humphries, M.J. *Biochem. Soc. Trans*. **2000**, *275*, 311-339.

26. Lee, J.-O.; Rieu, P.; Arnaout, M.A. Liddington, R. *Cell.* **1995**, *80*, 631–638.

27. Baldwin, E.T.; Sarver, R.W.; Bryant Jr., G.L.; Curry, K.A.; Fairbanks, M.B.; Finzel, B.C.; Heinrikson, R.L.; Horton, N.C.; Kelley, L.L. *Structure*. **1998**, *6*, 923-935.

28. Argraves, W. S.; Suzuki, S.; Arai, H.; Thompson, K.; Pierschbacher, M. D.; Ruoslahti, E. *J. Cell Biol.* **1987b**, *105*, 1183-1190.

29. Fitzgerald, L. A.; Steiner, B.; Rail, S. C.; Lo, S. S.; Philips, D. R. *J. Biol. Chem.* **1987**, 262, 3936-3939.

30. Hogervorst, F.; Kuikman, I.; von dam Borne, A. E. G. K.; Sonnenberg, A. *EMBO J.* **1990**, *9*, 765-770.

31. MacKrell, A. J.; Blumberg, B.; Haynes, S. R.; Fessler, J. H. *Proc. Natl. Acad. Sci. USA.* **1988**, *85*, 2633-2637.

32. Sheppard, D.; Rozzo, C.; Starr, L.; Quaranta, V.; Erie, D. J.; Pytela, R. *J. Biol. Chem.* **1990**, *265*, 11502-11507.

33. Springer, T.A. *Proc. Natl. Acad. Sci. USA.* **1997**, *94*, 65–72.

34. Lu, C.; Oxvig, C.; Springer, T.A. *J. Biol. Chem.* **1998**, *273*, 15138–15147.

35. Djaffar, I.; Chen, Y P.; Creminon, C.; Maclouf, J.; Cieutat, A.M.; Gayet, O.; Rosa, J.P. *Biochem. J.* **1994**, *300*, 69-74.

36. Zhou, L.; Lee, D. H. S.; Plescia, J.; Lau, C. Y.; Altieri, D. C. *J. Biol. Chem.* **1994**, *269*, 17075-17079.

37. Randi, A. M.; Hogg, N. *J. Biol. Chem.* **1994**, *269*, 12395 – 2398.

38. Tuckwell, D.; Calderwood, D. A.; Green, L. J.; Humphries, M. J. *J. Cell Sci.* **1995**, *108*, 1629 – 1637.

39. Calderwood, D. A.; Tuckwell, D. S.; Eble, J.; Kuhn, K.; Humphries, M. J. *J. Biol. Chem.* **1997**, *272*, 12311 - 12317

40. Emsley, J.; Knight, C. G.; Farndale, R. W.; Barnes, M. J.; Liddington, R. C. *Cell*. **2000**, *100*, 47 – 56.

41. Staunton, D. E.; Dustin, M. L.; Erickson, H. P.; Springer, T. A. *Cell*. **1990**, *61*, 243-254.

42. Michishita, M.; Videm, V.; Arnaout, M. A. *Cell*. **1993**, *72*, 857 – 867.

43. Bergelson, J. M.; Hemler, M. E. *Curr. Biol.* **1995**, *5*, 615–617.

44. D'Souza, S. E.; Haas, T.A.; Piotrowicz, R. S.; Byers-Ward, V.; McGrath, D. E.; Soule, H. R.; Cierniewski, C. S.; Plow, E. F.; Smith, J. W. *Cell*. **1994**, *79*, 659–667.

45. Hu, D. D.; Barbas III, C. F.; Smith, J. W. *J. Biol. Chem.* **1996**, *271*, 21745–21751.

46. Stanley, P.; Bates, P. A.; Harvey, J.; Bennett, R. I.; Hogg, N. *EMBO. J.* **1994**, *13*, 1790–1798.

47. Dickeson, S. K.; Walsh, J. J.; Santoro, S. A. *J. Biol. Chem.* **1997**, *272*, 7661–7668.

48. Bajt, M. L.; Goodman, T.; McGuire, S. L. *J. Biol. Chem.* **1995**, *270*, 94–98.

49. D'Souza, S. E.; Ginsberg, M. H.; Burke, T. A.; Plow, E. F. *J. Biol. Chem.* **1990**, *265*, 3440–3446.

50. Santoro, S. A.; Lawing Jr., W. J. *Cell*. **1987**, *48*, 867–873.

51. D'Souza, S. E.; Ginsberg, M. H.; Lam, S. C. T.; Plow, E. F. *J. Biol. Chem.* **1988**, *263*, 3943–3951.

52. Kamata, T.; Irie, A.; Tokuhira, M.; Takada, Y. *J. Biol. Chem.* **1996**, *271*, 18610–18615.

53. Tozer, E. C.; Baker, E. K.; Ginsberg, M. H.; Loftus, J. C. *Blood* . **1999**, *93*, 918–924.

54. Gulino, D.; Boudignon, C.; Zhang, L. Y.; Concord, E.; Rabiet, M. J.; Marguerie, G. *J. Biol. Chem.* **1992**, *267*, 1001–1007.

55. Xiong, J.P.; Stehle, T.; Zhang, R.; Joachimiak, A. Frech, M.; Goodman, S.L.; Arnaout, M.A. *Science*. **2002**, *296*, 151–155.

56. Charo, I.F.; Nannizzi, L.; Phillips, D.R.; Hsu, M.A.; Scarborough, R.M. *J. Biol. Chem.* **1991**, *266*, 1415-1421.

57. Larsen, B.S; McEwen, C.N. *Mass Spectrometry of Biological Materials*. Second edition. Marcel Dekker, Inc: New York, NY, 1998; pp 345-352

58. Kumar, C.C.; Malkowski, M.; Yin, Z.; Tanghetti, E.; Yaremko, B.; Nechuta, T.; Varner, J.; Liu, M.; Smith, E.M.; Neustadt, B.; Presta, M.; Armstrong, L. *Can. Res.* **2000**, *61*, 2232-2238.

59. Siuzdak, G. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91,* 11290.

60. Karas, M. Bachmann, D.; Bahr, U.; Hillenkamp, F. *Int. J. Mass Spectrom. Ion Proc*. **1987**, *78*, 53.

61. Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T. *Rapid Comm. Mass Spectrom.* **1988**, *2*, 151.

62. Hillenkamp, F.; Karas, M.; Beavis, R.C.; Chait, B.T. *Anal. Chem.* **1991**, *63*, 1193A.

63. Zenobi, R.; Knochenmuss, R. *Mass Spectrom. Rev.* **1998**, *17*, 337.

64. Pramanik, B.N; Ganguly, A.K; Gross M.L. *Applied Electrospray Mass Spectrometry*. Marcel Dekker, Inc: New York, NY, 2002; pp 361-370.

65. Herbert, C.G and Johnstone, R.A.W. *Mass Spectrometry Basics*. CRC press: New York, NY, 2003; pp55-60.

66. Kebarle, P.; Peschke, M. *Anal. Chimica Acta*. **2000**, *406*, 11–35

67. Schug, K.A. *Combin. Chem. High Throughput Screen*. **2007**, *10*, 301-316.

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

69. Mack, L.L.; Kralik, P.; Rheude, A.; Dole, M. J. *Chem. Phys.* **1970**, *52*, 4977.

70. Iribarne, J.V.; Thomson, B.A. *J. Chem. Phys.* **1976**, *64*, 2287.

71. Thomson, B.A.; Iribarne, J.V. *J. Chem. Phys.* **1979**, *71*, 4451.

72. Fenn, J.B.; Rosell, J.; Nohmi, T.; Banks Jr, F.J. *In Biochemical and Biotechnological Applications of Electrospray Ionization Mass Spectrometry. Snyder AP (ed.). American Chemical Society: Washington, DC,* **1995**; 60 - 80.

73. Fenn, J.B. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 524-535.

74. De la Mora, F. *Anal. Chim. Acta* **2000**, *406*, 93.

75. Wang, G.; Cole, R.B. *Anal. Chim. Acta* **2000**, *406*, 53.

76. Wang, G.; Cole, R.B. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 1050-1058.

77. Cole, R.B.; Harrata, A.K. *J. Am. Soc. Mass. Spectrom.* **1993**, *4*, 546.

78. Charles, L.; Pepin, D.; Gonnet, F.; Tabet, J.-C. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 1077-1084.

79. Wu, Z.; Gao, W.; Phelps, M.A.; Wu, D.; Miller, D.D.; Dalton, J.T. *Anal. Chem.* **2004**, *76*, 839-847.

80. Covey, T. *In Biochemical and Biotechnological Applications of Electrospray Ionization Mass Spectrometry. Snyder AP (ed.). American Chemical Society: Washington, DC,* **1995**; 21-59.

81. Wilm, M.; Mann, M. *Int. J. Mass Spectrom. Ion Processes.* **1994**, *136*, 167.

82. Wilm, M.; Mann, M. *Anal. Chem.* **1996**, *68*, 1.

83. Emmett, M.R. Caprioli, R.M. *J. Am. Soc.Mass Spectrom.* **1994**, *5*, 605.

84. Andren, P.E.; Emmett, M.R.; Caprioli, R.M. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 867

85. Collette, C.; De Pauw, E. *Rapid Commun. Mass Spectrom*. **1998**, *12*, 165.

86. Collette, C.; Drahos, E.L.; De Pauw, E.; Vekey, K. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 1673.

87. McLafferty, F.W. *Science*. **1981**, *214*, 280.

88. Schug, K.A.; Lindner, W. *J. Sep. Sci.* **2005**, *28*, 1932-1955.

89. Zehender, H.; Mayr, M. *Curr. Opin. Chem. Biol.* **2007**, *11*, 511-517.

90. Smith, R.D.; Light-Wahl, K.J. *Biol. Mass Spectrom.* **1993**, *22*, 493-501.

91. Smith, D.L.; Zhang, Z. *Mass Spectrom. Rev.* **1994**, *13*, 411-429.

92. Colton, R.; D'Agostino, A.; Traeger, J.C. *Mass Spectrom. Rev.* **1995**, *14*, 79-106.

93. Przybylski, M.; Glocker, M.O. *Angew. Chem. Int. Ed.* **1996**, *35*, 806-826.

94. Pramanik, B.N.; Bartner, P.L.; Mirza, U.A.; Liu, Y.-H.; Ganguly, A.K. *J. Mass Spectrom.* **1998**, *33*, 911-920.

95. Veenstra, T.D. *Biophys. Chem*. **1999**, *79*, 63-79.

96. Loo, J.A. *Int. J. Mass Spectrom.* **2000**, *200*, 175-186.

97. Nesatyy, V.J. *Int. J. Mass Spectrom.* **2002**, *221*, 147-161.

98. Finn, M.G. *Chirality.* **2002**, *14*, 534-540.

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

100. Tao, W.A.; Cooks, R.G. *Anal. Chem.* **2003**, *75*, 25A-31A.

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

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

103. Heck, A.J.R.; Jorgensen, T.J.D. *Int. J. Mass Spectrom.* **2004**, *236*, 11-23.

104. Di Tullio, A.; Reale, S.; De Angelis, F. *J. Mass Spectrom.* **2005**, *40*, 845-865.

105. Schug, K.; Fryčák, P.; Maier, N.; Lindner, W. *Anal. Chem*. **2005**, *77*, 3660-3670.

106. Lim, H.-K.; Hsieh, Y.L.; Ganem, B.; Henion, J. *J. Mass Spectrom.* **1995**, *30*, 708.

107. Greig, M.J.; Gaus, H.-J.; Cummins, L.L.; Sasmor, H.; Griffey, R.H. *J. Am. Chem. Soc.* **1995**, *117*, 10765.

108. Sannes-Lowery, K.A.; Griffey, R.H.; Hofstadler, S.A. *Anal. Biochem.* **2000**, *280*, 264.

109. Eckart, K.; Spiess, J. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 912.

110. Marky, L.A.; Breslauer, K.J. *Biopolymers.* **1987**, *26*, 1601.

111. Kempen, E.C.; Brodbelt, J.S. *Anal. Chem.* **2000**, *72*, 5411.

112. Frycak, P.; Schug, K.A. *Anal. Chem.* **2007**, *79*, 5407-5413.

113. Wortmann, A.; Kistler-Momotova, A.; Zenobi, R.; Heine, M.C.; Wilhelm, O.; Pratsinis, S.E. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 385-393.

114. Wang, H.; Agnes, G.R. *Anal. Chem.* **1999**, *71*, 3785.

115. Klotz, I.M. *Ligand-Receptor Energetics- A Guide for The Perplexed*, John Wiley& Sons Inc: New York, NY, 1997. pp 1- 65.

116. Schug, K.; Fryčák, P.; Maier, N.; Lindner, W. *Anal. Chem.* **2005**, *77*, 3660-3670

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

118. Gokel, G.W.; Goli, D.M.; Mingati, C.; Echegoyen, L. *J. Am. Chem. Soc.* **1983**, *105*, 6786.

119. Goff, C.M.; Matchetter, M.A.; Shabestary, N.; Khazaeli, S. *Polyhedron* **1996**, *15*, 3897.

120. Fändrich, M.; Tito, M.A.; Leroux, M.R.; Rostom, A.A.; Hartl, F.U.; Dobson, C.M.; Robinson, C.V. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 14151.

121. Shukla, A.K.; Futrell, J.H.; *J. Mass Spectrom.* **2000**, *35*, 1069.

122. Chen, Y.-L.; Campbell, J.M.; Collings, B.A.; Konermann, L.; Douglas, D.J. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 1003.

123. Gross, D.S.; Zhao, Y.; Williams, E.E. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 519.

124. Schnier, P.D.; Klassen, J.S.; Strittmatter, E.F.; Williams, E.R. *J. Am. Chem. Soc.* **1998**, *120*, 9605.

125. He, F.; Ramirez, J.; Garcia, B.A.; Lebrilla, C.B. *Int. J. Mass Spectrom.* **1999**, *182*, 261.

126. Rodgers, M.T.; Armentrout, P.B. *J. Am. Chem. Soc.* **2000**, *122*, 8548.

127. Rogniaux, H.; Van Dorsselaer, A.; Barth, P.; Biellmann, J.F.; Barbanton, J.; van Zandt, M.; Chevrier, B.; Howard, E.; Mitschler, A.; Potier, N.; Urzhumtseva, L.; Moras, D.; Podjarny, A. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 635.

128. Li, Y.-T.; Hsieh, Y.-L.; Henion, J.D.; Ocain, T.D.; Schiehser, G.A.; Ganem, B. *J. Am. Chem. Soc.* **1994**, *116*, 7487.

129. Penn, S.G.; He, F.; Green, M.K.; Lebrilla, C.B. *J. Am. Soc. Mass Spectrom.* **1997**, *8* 244.

130. Cooks, R.G.; Patrick, J.S.; Kotiaho, T.; McLuckey, S.A. *Mass Spectrom. Rev.* **1994**, *13*, 287.

131. Cooks, R.G.; Koskinen, J.T.; Thomas, P.D. *J. Mass Spectrom.* **1999**, *34*, 85.

132. Raji, M.A.; Frycak, P.; Beall, M.; Sakrout, M.; Ahn, J.-M.; Bao, Y.; Armstrong, D.W.; Schug, K.A. *Int. J. Mass Spectrom.* **2007**, *262*, 232-240.

133. Sharp, K.A.; Nicholls, A.; Friedman, R.; Honig, B. *Biochemistry*, **1991**, *30*, 9686-9697.

134. Harrison, A.G. *Mass Spectrom. Rev.* **1997**, *16*, 201–217.

135. Lias, S.G.; Bartmess, J.E. *NIST Web Book of Gas phase Thermochemistry.* http://webbook.nist.gov/chemistry/ion/

136. Perrin, D.D.; Dempsey, B.; Serjeant, E.P. *pK_a Prediction for Organic Acids and Bases* Chapman and Hall: New York, NY, 1981, pp. 19.

137. Haynie, D.T. *Biological Thermodynamics.* Cambridge University Press: Cambridge, UK, 2001, 343.

138. Kutner, M.H.; Neter, J.; Nachtsheim, C.J.; Li, W. *Applied Linear Statistical Models.* McGraw-Hill Education: Singapore, 2004; 5th International edition.

139. He, J.; Zelikovsky, A. *Bioinformatics* **2006**, *22*, 2558-2561.

140. Breiman, L.; Friedman, J.; Stone, C.J.; Olshen, R.A. *Classification and Regression Trees*. Chapman & Hall/CRC, 1984.

141. Morgan, E. *Chemometrics: Experimental Design;* John Wiley & Sons: Chichester, 1991, pp 81-150.

142. Rundlett, K.L.; Armstrong, D.W. *Electrophoresis.* **2001**, *22*, 1419 – 1427.

143. Schug, K.A.; Lindner, W. *International Journal of Mass Spectrometry* **2005**, *241*, 11–23.

144. Schug, K.A.; Lindner, W. *Chem. Rev.* **2005**, *105*, 67-113.

145. Loo, J.A.; Hu, P.; McConnell, P.; Mueller, T.W.; Sawyer, T.K.; Thanabal, V. *J. Am. Chem. Soc.* **1997**, *8*, 234-243.

146. Schalley, C.A.; Castellano, R.K.; Brody, M.S.; Rudkevich, D.M.; Siuzdak, G.; Rebek Jr., J. *J. Am. Chem. Soc.* **1999**, *121*, 4568-4579.

147. Zwiener, C.; Frimmel, F.H. *Anal. Bioanal. Chem.* **2004**, *378*, 851-861.

148. Giovannetti, R.; Bartocci, V.; Pucciarelli, F.; Ricciutelli, M. *Talanta* **2004**, *63*, 857–864.

149. Hofstadler, S.A.; Sannes-Lowery, K.A. *Nature Reviews Drug Discovery* **2006**, *5*, 585-595.

150. Hiraoka, K.; Murata, K.; Kudaka, I. *Journal of Mass Spectrometry Society of Japan.* **1995**, *43*, 127.

151. Sjoberg, Per J.R.; Bokman, F.C.; Bylund, D.; Markides, K.E. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 1002–1010.

152. Guevremont, R.; Siu, K.W.M.; Le Blanc, J.C.Y.; Berman, S.S. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 216-224.

153. Enke, C.G. *Anal. Chem.* **1997**, *69*, 4885-4893

154. Cech, N.B.; Enke, C.G. *Anal. Chem.* **2000**, *72*, 2717-2723.

155. Sherman, C.L.; Brodbelt, J.S. *Anal. Chem.* **2003**, *75*, 828-1836.

156. Zhou, S.; Cook, K.D. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 206–214.

157. Amad, M.H.; Cech, N.B.; Jackson, G.S.; Enke, C.G. *J. Mass Spectrom.* **2000**, *35*, 784–789.

158. Leize, E.; Jaffrezic, A.; Van Dorsselaer, A. *J. Mass Spectrom.* **1996**, *31*, 537-544.

159. Sakairi, M.; Yergey, A.L.; Siu, K.W.M.; Le Blanc, J.C.Y.; Guevremont, R.; Berman, S.S. *Anal. Chem.* **1991**, *63*, 1488-1490.

160. Tsodikov, O.V. *J. Comp. Chem.* **2002**, *23*, 600-609.

161. Brereton, R.G. *Chemometrics Data Analysis for the Laboratory and Chemical Plant*; John Wiley & Sons: Chichester, 2005, pp 15-31.

162. Brereton, R.G. *Applied Chemometrics for Scientists;* John Wiley & Sons: Chichester, 2005.

BIOGRAPHICAL INFORMATION

Misjudeen Raji was born in the small town of Ondo in Ondo state of Nigeria. He attended the prestigious St Joseph's College in Ondo, where he obtained his Senior School Certificate with Distinction in the Sciences. He proceeded to the University of Lagos, Nigeria, where he obtained a Bachelor of Science (Hons) degree in Industrial Chemistry. During his fourth year in the university, he embarked on a 6-month industrial internship program at the Nigerian National Petroleum Corporation, where he worked in the company's Health and Safety Department overseeing the local oil depots in the city of Lagos.

Following his first degree, he went on to carry out a mandatory 1-year National youth Service Corps program. During this program, he worked as a Management Trainee in the Finance and Tax department of PricewaterhouseCoopers. Upon completion of this program, he worked for nine months as a Quality Control Officer at Dangote Pasta Plant in Lagos, Nigeria. He later moved to the Nigerian Bottling Company (a subsidiary of the Coca Cola Company), where he worked as Quality Assurance Management Trainee for two months before he traveled to the US for graduate education.

Misjudeen Raji joined the University of Texas at Arlington in the Fall semester of 2002. He completed a non-thesis Masters of Science degree in Chemistry in December 2004 and later returned to the same department in Fall 2005 for his doctorate degree. During his doctoral degree, he worked under the tutelage of Dr Kevin Schug on the investigation of integrin-RGD molecular recognition using ESI-MS. He later went on to pursue post-doctoral research fellowship under the auspices of Dr Michael Olivier at the Medical College of Wisconsin.