INVESTIGATION ON GAS-PHASE ION STRUCTURES

OF BIOMOLECULES USING

ION MOBILITY-MASS SPECTROMETRY

A Dissertation

by

LEI TAO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2010

Major Subject: Chemistry

INVESTIGATION ON GAS-PHASE ION STRUCTURES

OF BIOMOLECULES USING

ION MOBILITY-MASS SPECTROMETRY

A Dissertation

by

LEI TAO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	David H. Russell
Committee Members,	Emile A. Schweikert
	Gyula Vigh
	Arul Jayaraman
Head of Department,	David H. Russell

May 2010

Major Subject: Chemistry

ABSTRACT

Investigation on Gas-phase Ion Structures of Biomolecules

Using Ion Mobility-mass Spectrometry. (May 2010)

Lei Tao, B.A., Fudan University, Shanghai, China;

M.S., The University of Akron

Chair of Advisory Committee: Dr. David H. Russell

IM-MS is a 2-D technique which provides separations based on ion shape (ion-neutral collision cross-section, Ω) and mass (m/z ratio). Ion structures can be deduced from the measured collision cross-section (Ω_{meas}) by calculating the collision cross-sections (Ω_{calc}) of candidates generated by molecular dynamics (MD) and compared with the experiment results.

A database of Ωs for singly-charged peptide ions is presented. Standard proteins are digested using different enzymes (trypsin, chymotrypsin and pepsin), resulting in peptides that differ in amino acid composition. The majority (63%) of the peptide ion correlates well with the globular structures, but some exhibit Ωs that are significantly larger or smaller than the average correlation. Of the peptide ions having larger Ωs, approximately 71% are derived from trypsin digestion, and most of the peptide ions that have smaller Ωs are derived from pepsin digestion (90%).

We use computational simulations and clustering methods to assign backbone conformations for singly-protonated ions of the model peptide (NH_2-Met-Ile-Phe-Ala-

Gly-Ile-Lys-COOH) formed by both MALDI and ESI and compare the structures of MIFAGIK derivatives to test the 'sensitivity' of the cluster analysis method. Cluster analysis suggests that [MIFAGIK + H]$^+$ ions formed by MALDI have a predominantly turn structure even though the low energy ions prefer partial helical conformers. Although the ions formed by ESI have Ωs that are different from those formed by MALDI, the results of cluster analysis indicate that the ions backbone structures are similar. Chemical modifications (N-acetyl, methylester, as well as addition of Boc or Fmoc groups) of MIFAGIK, alter the distribution of various conformers. The most dramatic changes are observed for the [M + Na]$^+$ ion, which show a strong preference for random coil conformers, owing to the strong solvation by the backbone amide groups.

Ω_{meas} of oligodeoxynucleotides in different length have been measured in both positive and negative modes. For a given molecular weight and charge state, Ω_{meas} of the oligodeoxynucleotide, ions are smaller than those of the peptides, indicating their different packing efficiency. A novel generalized non-Boltzman sampling MD has been utilized to investigate the gas-phase ion conformations of dGGATC based on the free energy values. Theory predicts only one low-energy conformer for the zwitterionic form of dGGATC$^-$ while dGGATC$^+$ ions have several stable conformers in both canonical and zwitterionic form in the gas phase, in good agreement with the experiment.

DEDICATION

This dissertation is dedicated my husband, Chang Zhang, my parents, Fenggang Tao and Lihua Li, my sister, Hua Tao, and her family for their love and support during my graduate career.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE   Page

FIGURE   Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

Understanding the molecular basis of biomolecule structure and function is currently a major focus of biological research. The predominant techniques for determining atomic-resolution structures of biomolecules are X-ray crystallography [1] and Nuclear Magnetic Resonance (NMR) [2]. However, high throughput analysis of complex biological mixtures obtained by using these techniques is underdeveloped, and serves as major bottleneck.

A major advantage of mass spectrometry is the ability to characterize biomolecules rapidly at very high sensitivity, and with high accuracy. The development of electrospray ionization (ESI) [3] and matrix-assisted laser desorption/ionization (MALDI) [4] has resulted in rapid adavance in biomolecular analyses at physiological levels. MALDI and ESI are both soft ionization techniques and are sufficiently gentle to generate ions of intact biomacromolecules with molecular masses above a few thousand Daltons (Da) even beyond 100 kDa [5, 6] without fragmentation. In addition to the accurate determination of molecular weights, these desorption/ionization MS methods have several established applications to primary structure analysis of proteins, such as full or partial sequence determination, characterization of intracellular proteolytic processing pathways and the identification of covalent post-translational structure modification with the combination of specific chemical (e.g., enzymatic) modification or

_____

This dissertation follows the style of *Journal of the American Society for Mass Spectrometry*.

collision-induced fragmentation technique [7, 8]. The emphasis of mass spectrometry based biological chemistry is shifting from compound identification to structural studies of large biomolecules and biomolecule complexes [9-15], including membrane proteins [16]. Recently, applications designed to obtain additional information of higher-order structures, such as secondary, tertiary, and quaternary structures, to study structure-function and even to probe specific non-covalent interactions between peptides and proteins, proteins and their substrates, and nucleic acids are becoming the new topics of mass spectrometry [5, 17, 18].

However, the investigations of biomolecules using mass spectrometry are carried out in the gas phase and while the aqueous phase ($\varepsilon=78$) presents the 'natural' environment for biomolecules. As for any gas phase methods, the challenges for mass spectrometry are in establishing the relevance of its results to the solution state conformations. McLafferty and coworkers used gas phase hydrogen/deuterium (H/D) exchange to observe the process of the folding and unfolding of cytochrome *c* ions in the gas phase [19], which is analogous to the way pH changes modulate folding in solution phase. The study showed both reversible and irreversible thermal transitions between several different conformational states. However, it is still questionable whether these transitions are the same state as the folding in the solution phase since the number of protons available for exchange for the gas phase structure might differ from that of the native structure in solution. More recently, Oomens *et. al.* have reported the infrared spectroscopic investigation of cytochrome c in the gas phase. Infrared spectra are obtained by multiple photon dissociation spectroscopy of the protein isolated in a

Fourier Transform Ion Cyclotron Resonance (FT-ICR) mass spectrometer. The spectra contain clearly resolvable bands in the amide I (C=O stretching modes) and amide II (N–H bending modes) spectral regions, slight shifted from those of cytochrome c in solution phase. The band positions are suggestive of a mostly α-helical structure of the protein and their widths are comparable to those in solution, suggesting the gas phase structure is an intrinsic property of the protein which is maintained in solution.

As known, the forces which control biomolecular structure in the native state are consist of the intramolecular interactions within the biomolecule and intermolecular hydration interactions between the biomolecule and its solvent, including hydrogen bonding [20], hydrophobic forces [21], and packing [22]. Each of these plays a crucial role for the structure and function of biomolecules. H-bonding is perhaps the most important interaction governing biomolecular structure, folding, binding, enzyme catalysis, and other properties [23]. The basic secondary structural elements in protein structure such as α-helical, β-sheet, γ-turn, π-helix, etc., are stabilized by H-bonding interactions [20]. The double helical structure of DNA and RNA are dominated by hydrogen bonds and stacking interactions, facilitating molecular recognition via replication processes and protein synthesis [24]. From the point of view of H-bonding, without the alternative of H-bonding to the solvent, the secondary structures of biomolecules should be even more thermodynamically stable in the gas phase. Indeed, it is of fundamental interest to investigate the biomolecular structure in the solvent-free gas phase. The gas-phase study of biomolecular structure provides several unique advantages over solution methods: i) it is possible to detect and characterize the

molecules too large for study via solution techniques such as NMR or some biological important molecules such as membrane proteins which are difficult to crystallize; ii) it also provides information on the intrinsic intramolecular interactions in the biomolecule without any concern of solvation effect and can thus help elucidate the driving forces that influence protein conformation and hence aid in solving the protein-folding problem. The techniques have been implemented for studying gas-phase biomolecules include chemical methods like H/D exchange and physical methods based on measurements of size, *i.e.*, ion mobility.

In the gas-phase H/D exchange experiment, the ions of interest are trapped for long times while exposed to a small amount of D_2O. The number of labile hydrogen can be obtained from the mass spectrum and the exchange rate can thus be measured. As the hydrogen atoms are involved in formation of hydrogen bonds in secondary structural elements, *i.e.*, α-helices and β-sheets; the exchange rate is a reflection of structure and structural stability. McLafferty and coworkers have reported several H/D exchange studies of protonated cytochrome *c* using Fourier transform ion cyclotron resonance (FT-ICR) [19, 25, 26]. These studies indicate that gas-phase folding involves not just two states but several different populations of protein molecules that display varying degrees of proton exchangeability and different degrees of openness. The gas-phase H/D exchange has also been studied in amino acids and small peptides [27-29]. However, the results suggest that are even for simple systems the interpretation of gas-phase H/D exchange data is not straightforward.

In the mobility measurement, ions are separated based on their ion-neutral collision cross-section with buffer gas under a weak electric field, and the collision cross-section data can be transformed to structural information of ions using molecular dynamics methods. The use of ion mobility spectrometry (IMS) as an important analytical technique for detecting and identifying volatile compounds dates back to the early 1980's [30-33]. Indeed, IMS is much more than a separation device. In the 1990s, several groups combined IMS with high-performance mass spectrometers to provide accurate mass measurements of ions exiting the IM drift cell [34]. More recently, IM-MS instruments that operate as tandem mass spectrometry instruments (IM-MS/MS) have also been developed [35-42]. Potential advantages of IM-MS and IM-MS/MS for proteomics research are post-ionization separation, which facilitates direct analysis of complex mixtures, increased sample throughput afforded by rapid data acquisition (μs-ms), and reduction of chemical noise by separation of molecular classes owing to differences in intrinsic gas-phase packing efficiencies of the ions [43, 44]. We refer to the separation of molecular classes in the mobility-mass dimension in terms of "conformation space," which is realized by plotting collision cross-section vs. *m/z* ratio of the ion [34]. The structural information derived from 2D conformation space afforded by IM-MS is potentially well-suited to both high throughput applications and complex biological samples.

The groups of Bowers and Jarrold are the pioneers to start the area of ion mobility research to obtain structural information of biomolecules in the gas phase [45-50]. Such IM-MS applications typically involve extensive theoretical work including *ab*

initio or density functional theory, and molecular mechanics or molecular dynamics methods to generate reasonable model structures for the molecular system studied experimentally. Clemmer and co-workers have compiled a database of peptide ion collision cross-sections and suggested that the majority of small (< 1500 m/z), singly-charged, tryptic peptides do not exhibit anhydrous secondary structure [51]. That is, most peptide ions assume a compact, charge-solvated (globular) structure in the gas phase [52, 53] and appear on a single trendline in 2-D mobility-m/z plots, i.e., plots of arrival-time distribution (ATD) or Ω vs. m/z. However, a small number of ion signals deviate (>3 to ~20%) from the expected trendline, and non-peptidic ion signals appear on separate, compound class specific trendlines [34, 54]. Ruotolo *et al.* showed that gas-phase $[M + H]^+$ ions of LLGNVLVVVLAR (derived from bovine hemoglobin) prefer extended (helical) structure(s) resulting in a larger Ω_{meas} than random coil structures having the same or similar m/z values [55, 56]. Peptides such as $(AAKAA)_n$ (n = 3-6) and $(AEAAKA)_n$ (n = 2-7) also show distinct structures, helical, and random coil depending on length and charge site, while some post-translational modified (PTM) peptide ions (phosphopeptides) tend to pack more tightly than the unmodified protonated peptide ions owing to intra-molecular charge-solvation and/or formation of salt-bridged type structures [57, 58]. In addition, our group has used chemical derivatization of the N-terminal and internal basic (acetylation) and acidic (methylation) residues (lysine and glutamic acid) to show that the helical propensity of a given peptide can be increased by reducing the number of 'salt-bridge' intramolecular interactions [59].

The challenge for structure IM-MS is to design gas-phase experiments that critically evaluate the structural assignments [60-62]. The research presented here is aimed to construct a bridge that links the experimental domain of IM-MS and the MD simulations and bioinformatic tools that are essential to interpretation of the data. In this study, we establish a database of high confidence level peptide ion collision cross-sections values which can be used to define the regions of conformation space occupied by peptide $[M + H]^+$ ions over a range of values typically encountered in "bottom-up" MS based proteomics. With extensive computational simulations and novel clustering methods, detail structural information for peptides of interest is obtained. The comparison of structural difference between gas-phase peptides generated by ESI and MALDI and the effect of the amino acid sequence and charge location on helix formation of peptide ions in the gas phase will be discussed. A series of oligonucleotide of different length will be tested by MALDI-IM-TOFMS to define the conformational space of oligonucleotide. With a novel non-Boltzman sampling molecular dynamics (MD) on model oliogonucleotide ions, a more accurate description of the gas phase conformational space of oligonucleotide ions can be achieved, which involving the thermodynamically accessible structures that correspond to real experimental condition.

CHAPTER II

A COLLISION CROSS-SECTION DATABASE OF

SINGLY-CHARGED PEPTIDE IONS*

Introduction

Ion mobility (IM) spectrometry separates ions on the basis of ion-neutral collision cross-section or apparent surface area, and several groups have combined IM spectrometry with high-performance mass spectrometers to provide accurate mass measurements of ions exiting the IM drift cell [34]. More recently, IM-MS instruments that operate as tandem mass spectrometry instruments (IM-MS/MS) have also been developed [35-42]. Potential advantages of IM-MS and IM-MS/MS for proteomics research are post-ionization separation, which facilitates direct analysis of complex mixtures, increased sample throughput afforded by rapid data acquisition (μs-ms), and reduction of chemical noise by separation of molecular classes owing to differences in intrinsic gas-phase packing efficiencies of the ions [43, 44]. We refer to the separation of molecular classes in the mobility-mass dimension in terms of "conformation space," which is realized by plotting collision cross-section vs. *m/z* ratio of the ion [34].

Several research groups have focused on developing IM-MS for biophysical studies of peptides and proteins [34, 52, 55, 56, 63, 64]. Our laboratory as well as Clemmer's has focused considerable attention on developing applications of IM-MS for

_____

proteomics, *i.e.,* high throughput analysis of peptides/proteins, and most of this work has utilized "bottom-up" proteomics or the identification of proteins by enzymatic digestion of the protein followed by MS analysis of the peptide fragments [34, 65]. Clemmer and co-workers have compiled a database of peptide ion collision cross-sections and suggested that the majority of small (< 1500 m/z), singly-charged, tryptic peptides do not exhibit anhydrous secondary structure [51]. That is, most peptide ions assume a compact, charge-solvated (globular) structure in the gas phase [52, 53]; however, we showed that a small number of gas-phase, tryptic peptides prefer extended (helical) structure in the gas phase [55, 56]. In some cases post-translational modification results in smaller than predicted structures from the average mobility-mass correlation, which likely results from intra-molecular charge-solvation and/or formation of salt-bridged type structures. We are interested in evaluating IM-MS as a screening method for identifying peptides which exhibit secondary structure or are post-translationally modified [34], which could provide an added data dimension for protein identification as well as unraveling the molecular parameters that most strongly influence higher order structure in the absence of water.

A primary challenge associated with developing IM-MS for proteomics scale protein identification is the extensive molecular simulations required for cross-correlating collision cross-sections with ion structure. The goal of this study is to establish a database of high confidence level peptide ion collision cross-sections values which can be used to define the regions of conformation space occupied by peptide $[M + H]^+$ ions over a range of values typically encountered in "bottom-up" MS based

proteomics. Our working definition of "high confidence level" includes correlation of peptide ion collision cross-sections with accurate mass measurements and sequence assignments based on tandem MS data.

Experimental

All proteins (bovine β-casein (M_r 25,091 Da), bovine serum albumin (M_r 69,248 Da), horse cytochrome c (M_r 11,694Da), horse hemoglobin α-chain (M_r 15,105 Da), horse hemoglobin β-chain (M_r 15,998 Da), horse myoglobin (M_r 16,941 Da), chicken egg ovalbumin (M_r 42,722 Da), chicken egg lysozyme (M_r 16,228 Da), rabbit aldolase (M_r 39,187 Da), yeast enolase (M_r 46,642 Da) were purchased from Sigma (St. Louis, MO) and used without further purification.

Trypsin and chymotrypsin digestions were performed using standard protocols as described elsewhere [66]. Briefly, proteins were thermally-denatured at 90 °C for 20 min, and then they were enzymatically-digested with sequencing-grade trypsin (Promega, Madison, WI) or chymotrypsin (Sigma) in a 1: 40 (w/w) enzyme to analyte ratio at 37 °C for ca. 20 h. Proteins were digested with pepsin (Sigma) using standard protocols. The pH of the protein solution was lowered to ca. 2 by adding 0.1M HCl, mixed with pepsin (1:20 protein to enzyme ratio (w/w)) and incubated at 37 ºC for 2 h. To quench the reaction, the pH of the digest was neutralized with 0.1M NH_4OH.

MALDI was performed by mixing the protein digests with re-crystallized α-cyano-4-hydroxycinnamic acid (Sigma) in a ca. 2000:1 matrix to analyte molar ratio.

The mixture was directly deposited as a dried droplet on a stainless steel MALDI sample plate.

Peptide amino acid sequences were initially assigned based on *in silico* protein digests generated by "PeptideMass" on the ExPASy website [67] and high resolution MALDI-TOF mass analysis (PerSeptive Biosystems, Voyager DE STR). Sequence assignments for selected peptides were confirmed using MALDI-based tandem mass spectrometry (Applied Biosystems 4700 Proteomics analyzer).

The MALDI-IM-TOFMS instrument used for these studies was constructed in collaboration with Ionwerks Inc. (Houston, TX), and the basic instrument design is similar to those previously described [68]. Briefly, MALDI was performed using high repetition rate-frequency tripled (355 nm) Nd:YAG laser (CrystaLaser, Reno, NV) operated at a pulse rate of 200-400 Hz. The resulting ions were introduced into a 15 cm drift tube maintained at a pressure of ca. 2.5 Torr He at room temperature and using IM field strengths ranging from 30 to 48 V/cm·Torr. Ions were separated based on mobility with a resolution of ca. 30 to 50. The ions eluting the drift cell were extracted into an orthogonal reflectron-TOF for mass analysis (resolution, ca. 2000 - 4000). The mass spectrometer was externally calibrated using two-point calibration of the radical cations (+1) of C_{60} (M_r = 720) and C_{70} (M_r = 840) (Sigma) [35]. The 2D IM-MS data was acquired and processed by using custom software (Ionwerks, Inc.).

All spectra were acquired under "low-field" conditions [69-71] using five different IM field strengths to accurately estimate the mass-dependent drift time correction, t_o ($t_{measured} - t_o = t_d$), which represents time the ion spends outside of the IM

drift cell. Collision cross-sections were calculated using the empirical drift times (t_d) and the hard sphere approximation [69]:

$$\Omega_{avg} = \frac{(18\pi)^{\frac{1}{2}}}{16} \frac{ze}{(k_b T)^{\frac{1}{2}}} \left(\frac{1}{m_I} + \frac{1}{m_B} \right)^{\frac{1}{2}} \frac{t_d E}{L} \frac{760}{P} \frac{T}{273.2} \frac{1}{N_0}$$ (1)

where z is the charge of the ion, e is elementary charge, N_0 is the number density of the drift gas at STP, k_b is Boltzmann's constant, m_I is the mass of the ion, m_B is the mass of buffer gas, E is the IM electric field strength, L is the drift tube length, P is the buffer gas pressure and T is the system temperature. To evaluate the accuracy of our collision cross-section measurements, we regularly measured the collision cross-section of bradykinin [M + H]$^+$ ions. The measured values of bradykinin [M+H]$^+$ ions are always within 2% of 242 Å^2, consistent with previously reported values for bradykinin [M+H]$^+$ ions [72].

Molecular dynamics calculations were performed as described elsewhere [56, 73]. Briefly, simulated annealing was performed with Cerius2 (Accelrys, San Diego, CA) using the consistent force field (CFF 1.02). During the simulation, the starting structure was gradually heated from 300 K to 1000 K over a 280 ps cycle. After each annealing cycle, a minimized structure was produced. Annealing cycles were repeated 300 times for each starting structure. The collision cross-section of each structure was calculated using the trajectory method in the MOBCAL software [74]. The representative structures are the lowest energy structures within 2 % of the experimental collision cross-section.

Results and Discussion

Ion-neutral collision cross-section database of singly-charged peptide ions

The objective of this study is to build a database of IM-MS data for singly-charged peptide ions. The database is composed of a diverse set peptide sequences which were generated by digesting nine standard proteins with three proteolytic enzymes (trypsin, chymotrypsin and pepsin). Trypsin cleaves proteins to yield peptides having Lys and Arg at the C-terminus, chymotrypsin cleaves on the C-terminal side of Trp, Tyr, Met, Leu or Phe, and pepsin cleaves C-terminal to Ala, Leu, Phe, or Tyr (Table 1). Because each of these enzymes has very different cleavage sites, we were able to produce a diverse mixture of peptides that can be used to examine peptide "conformation space", *i.e.,* 2D plot of ion mobility collision cross-section vs. *m/z* values.

The peptides examined in study were characterized by three methods: (i) accurate *m/z* values of all peptides were obtained by using high resolution MALDI-MS and compared to a peptide fragment list generated *in silico* [67], (ii) peptide assignments were confirmed using MALDI-TOF/TOFMS, and (iii) the peptide ion collision cross-sections were measured using the MALDI-IMMS. The data from these experiments are limited to singly-charged peptide ions. A total of 607 peptides from digestion are identified by MALDI-IM-TOFMS (shown in Table 2): 152 tryptic peptides, 295 chymotryptic peptides, and 139 peptic peptides, and a small number of redundant peptides. That is, 21 of the peptides generated by chymotrypsin were also found in the pepsin digests.

Table 1. Cleavage sites of different enzymes.

Enzyme	Cleaves at:	Except if
Trypsin	C-term to K/R	before P
Chymotrypsin	C-term to F/L/M/W/Y	before P
Pepsin (pH > 2)	C-term to A/E/F/L/Q/W/Y	

Table 2. Protein digestion results using three enzymes identified by MALDI-IM-TOFMS.

Protein	MW	*Trypsin digestion*			*Chymotrypsin digestion*			*Pepsin digestion*		
		# Matched Peptides	Mass Range	Coverage (%)	# Matched Peptides	Mass Range	Coverage (%)	# Matched Peptides	Mass Range	Coverage (%)
ALBU_BOVIN	69248	27	584 ~ 2045	35	64	513 ~ 2803	63	26	729 ~ 5125	17
CYC_HORSE	11694	14	762 ~ 2209	87	17	619 ~ 2511	88	13	584 ~ 3111	77
HBA_HORSE	15105	5	1087 ~ 3077	50	30	522 ~ 2546	97	14	582 ~ 3781	89
HBB_HORSE	15998	8	1126 ~ 2326	69	31	515 ~ 1827	86	22	544 ~ 4193	90
LYSC_CHICK	16228	14	517 ~ 3136	77	28	577 ~ 2725	94	17	903 ~ 5050	98
OVAL_CHICK	42722	23	580 ~ 2460	59	49	519 ~ 3074	79	35	516 ~ 3055	54
ALDO_RABIT	39187	20	584 ~ 2272	48	42	529 ~ 3351	83	17	692 ~ 1935	38
CASB_BOVIN	25091	15	517 ~ 2925	29	23	603 ~ 3616	77	15	675 ~ 3205	63
MYG_HORSE	16941	17	631 ~ 2601	97	32	548 ~ 2214	100			
ENO1_YEAST	46642	26	508 ~ 3737	74						
Total number of peptides resolved in MALDI-IM-TOF:		169			316			159		

The ion-neutral collision cross-sections of these peptides are listed in Appendix A and the database is also available at the link listed below: http://www.chem.tamu.edu/rgroup/russell/cross_section_database.pdf. The peptides included in the database do not include post-translationally modified peptides, with the exception of three oxidized peptides.

Figure 1 contains a plot of conformation space for the entire peptide database. Most peptide signals are clustered along a nonlinear "trendline," which we will refer to as the "average mobility-mass correlation" [34]. The solid line in Figure 1 is a third-order polynomial fit, R^2 = 0.988) to the individual data points. The majority (63%) of peptide signals correlate quite well (within 3%) to the average, globular mobility-mass correlation (Table 3). In our previous work we used linear regression of the data to describe the correlation between collision cross-section and *m/z* [58, 75], but linear regression analysis is only valid over a small range of *m/z* values, owing to the effective surface area-to-*m/z* relationships for peptide conformation space. Thus, as the diameter (d) of a spherical (globular) ion increases, its volume (d^3, proportional to mass) increases more quickly than the surface area (d^2), which results in a nonlinear correlation between collision cross-section and *m/z*. The average mobility-mass correlation derived from the data shown in Figure 1 is consistent with the previous database generated by Clemmer and colleagues (a dataset of 420 singly-charged peptides generated by electrospray ionization and analyzed by IM-TOFMS) [51].
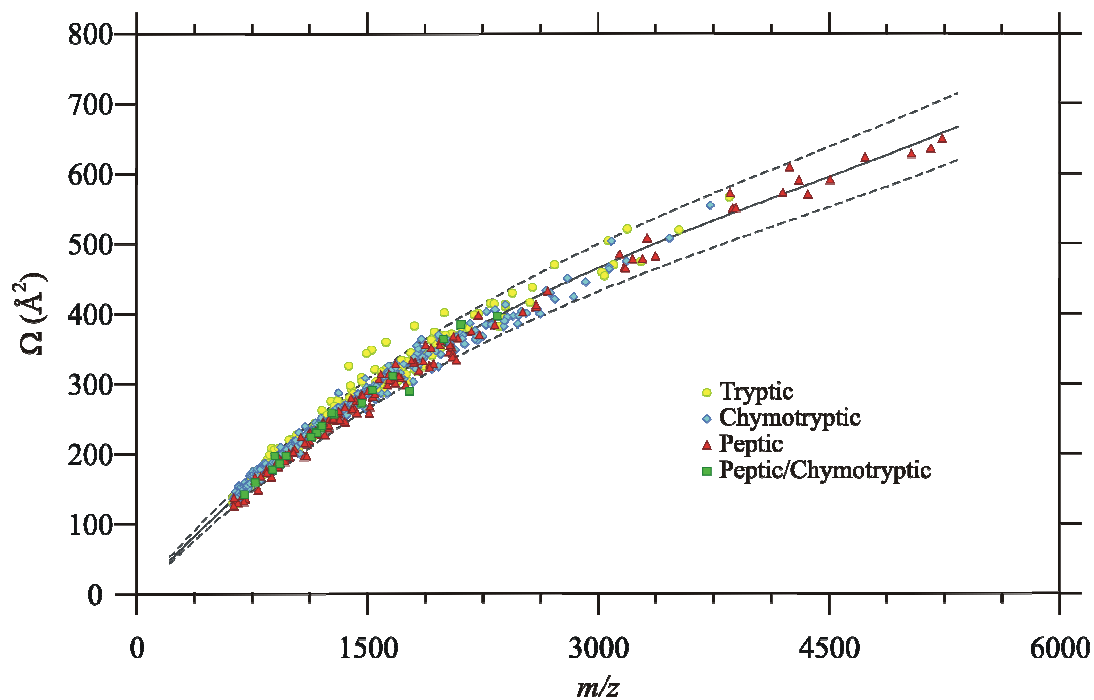
Figure 1. Plot of ion-neutral collision cross-sections vs. *m/z* for 607 [M + H]$^+$ peptide ions. The solid line corresponds to a third-order polynomial fit to the data (R^2 = 0.988) and corresponds to the average mobility-mass correlation (y = $2.81*10^{-9}x^3$ - $3.55*10^{-5}x^2$ + $2.32*10^{-1}x$ + 41.91, R^2 = 0.988). The dashed lines represent ± 7% deviation from the fit.

Table 3. Compilation of selected data for peptide ions which fall within 3 % of the average mobility-mass correlation.

Enz.	Protein	Position	Sequence	m/z	Ω (Å²)	% Dev.
C	ALBU_BOVIN	147-150	CDEF	513.17	152 ± 2	-0.1
C	HBB_HORSE	38-41	TQRF	551.29	160 ± 3	0
T	ENO1_YEAST	234-240	AAGHDGK	655.32	180 ± 3	0.1
C	HBB_HORSE	142-146	AHKYH	655.33	180 ± 1	0
T	MYG_HORSE	57-62	ASEDLK	662.34	181 ± 3	0
P	ALDOA_RABIT	357-363	FISNHAY*	851.4	216 ± 2	0.1
C	CYC_HORSE	1--10	GDVEKGKKIF	1120.64	262 ± 3	0.1
P	HBB_HORSE	33-41	VVYPWTQRF*	1195.63	273 ± 1	0
C	CYC_HORSE	37-48	GRKTGQAPGFTY	1282.65	287 ± 3	-0.1
C	ALDOA_RABIT	31-43	AADESTGSIAKRL*	1318.7	292 ± 2	-0.1
C	ALBU_BOVIN	435-446	TRKVPQVSTPTL	1326.77	294 ± 2	0
C	OVAL_CHICK	19-29	KVHHANENIFY	1371.68	301 ± 2	0.1
T	MYG_HORSE	119-133	HPGDFGADAQGAMTK	1502.67	320 ± 4	0
T	HBA_HORSE	17-31	VGGHAGEYGAEALER	1515.72	322 ± 2	0.1
C	HBA_HORSE	35-48	GFPTTKTYFPHFDL	1670.82	344 ± 2	0.1
T	LYSC_CHICK	116-130	IVSDGNGMNAWAWR	1675.8	344 ± 4	0
C	HBB_HORSE	13-28	ALWDKVNEEVGGEAL	1758.85	355 ± 3	0
T	MYG_HORSE	1--16	GLSDGEWQQVLNVWGK*	1815.9	363 ± 4	0.1
T	ALBU_BOVIN	508-523	RPCSALTPDETYVPK	1823.9	364 ± 1	0
P	HBB_HORSE	63-85	HGKKVLHSFGEGVHHLDNLK GTF	2557.34	450 ±2	0

Note: the table list the enzyme used for proteolysis (Enz., T = trypsin, C = chymotrypsin, P = pepsin, the UniProtKB protein identifier, position of peptide in the protein, amino acid sequence, monoisotopic mass-to-charge ratio (m/z), collision cross-section (Ω), and percent deviation from the average mobility-mass correlation (% Dev.). Collision cross-sections and reduced mobilities are reported as the average ± 1σ for 5 replicate measurements. (*) Indicates peptide sequences were confirmed by tandem MS.

There are 37 (out of 607) peptide ions that have collision cross-sections that are larger than 7% from that expected for globular ions, and most (71%) of these are tryptic peptides (C-terminal Lys or Arg residues). Table 4 lists twenty of them with the largest deviation from the globular correlation. Jarrold showed that positively-charged peptide ions with basic residues near the C-terminus have relatively high helical propensities [52, 53, 76], and we have found similar trends for a series of model peptides specifically designed to examine the effects of C-term charge carriers on gas-phase ion structure [59]. Thus, we suggest that peptide ions having larger than predicted collision cross-sections correspond to extended conformations and can be good candidates for helical structures.

There are ten peptides that exhibit collision cross-sections which are ca. 7 % smaller (more compact structures) than the average (Table 5). The majority of these peptides are the products of pepsin digestion. We previously showed that gas-phase phosphorylated peptide ions exhibit compact structures [34, 58, 77]. Although the peptides which fall below the average mobility-mass correlation in our database are not modified, many of them have multiple acidic and polar residues which could lead to formation of salt-bridged structures and/or H-bonding between acidic and basic side chains, which could significantly reduce collision cross-sections [73, 78]. Furthermore, the average number of polar residues increases across the classes of peptides listed in Tables 1-3; for peptides that fall on or near the globular mobility-mass correlation (Table 1), approximately 48 % are polar, whereas only 34 % of the residues of the peptides which occur above the globular trendline (Table 2) are polar. The compact peptide ions which fall below the globular mobility-mass correlation (Table 3) are 59 % polar.

Table 4. Compilation of selected data for peptide ions with collision cross-section that deviate by more than +7% from the average mobility-mass correlation.

Enz.	Protein	Position	Sequence	m/z	Ω (Å²)	% Dev.
C	MYG_HORSE	30-33	IRLF	548.36	173 ± 1	8.6
T	MYG_HORSE	134-139	ALELFR*	748.44	215 ± 2	9.3
T	ENO1_YEAST	409-414	LNQLLR	756.47	215 ± 2	8.5
T	CYC_HORSE	8--13	KIFVQK	762.49	225 ± 3	12.9
T	CYC_HORSE	80-86	MIFAGIK*	779.45	223 ± 4	10.2
T	CYC_HORSE	73-79	KYIPGTK	806.48	225 ± 4	8.6
T	HBA_HORSE	91-99	LRVDPVNFK	1087.63	280 ± 2	9.4
T	HBB_HORSE	133-144	VVAGVANALAHK	1149.67	292 ± 2	9.8
C	HBB_HORSE	104-114	RLLGNVLVVVL	1194.79	304 ± 4	11.3
C	HBA_HORSE	126-136	DKFLSSVSTVL	1195.66	304 ± 2	11.2
T	HBB_HORSE	105-116	LLGNVLVVVLAR*	1265.83	343 ± 2	20.6
T	HBB_HORSE	31-40	LLVVYPWTQR*	1274.73	315 ± 2	10
T	OVAL_CHICK	370-381	HIATNAVLFFGR*	1345.74	327 ± 3	10.1
T	MYG_HORSE	64-77	HGTVVLTALGGILK*	1378.84	361 ± 2	19.6
T	CASB_BOVIN	206-217	LLYQEPVLGPVR	1383.8	330 ± 2	9
T	ENO1_YEAST	105-119	LGANAILGVSLAASR*	1412.82	366 ± 2	19.2
T	CYC_HORSE	26-38	HKTGPNLHGLFGR*	1433.78	338 ± 3	8.9
T	MYG_HORSE	64-78	HGTVVLTALGGILKK*	1506.94	378 ± 5	17.8
T	ALDOA_RABIT	243-257	YSHEEIAMATVTALR*	1691.84	400 ± 1	15.5
T	MYG_HORSE	103-118	YLEFISDAIIHVLHSK*	1885.02	419 ± 4	12.7

Note: the table list the enzyme used for proteolysis (Enz., T = trypsin, C = chymotrypsin, P = pepsin, the UniProtKB protein identifier, position of peptide in the protein, amino acid sequence, monoisotopic mass-to-charge ratio (m/z), collision cross-section (Ω), and percent deviation from the average mobility-mass correlation (% Dev.). Collision cross-sections and reduced mobilities are reported as the average ± 1σ for 5 replicate measurements. (*) Indicates peptide sequences were confirmed by tandem MS.

Table 5. Compilation of selected data for peptide ions that deviate by more than -7% from the average mobility-mass correlation. (Notations are as same as table1)

Enz.	Protein	Position	Sequence	m/z	Ω (Å^2)	% Dev.
P	CYC_HORSE	44-48	PGFTY	584.27	149 ± 2	-10.2
P	ALDOA_RABIT	54-57	NRRF	592.33	153 ± 1	-8.8
P	CASB_BOVIN	16-20	RELEE*	675.33	166 ± 1	-9.7
P	OVAL_CHICK	84-89	RDILNQ	758.42	184 ± 1	-7.3
P	OVAL_CHICK	112-119	PILPEYLQ	972.54	213 ± 1	-9.8
P	OVAL_CHICK	135-143	QTAADQARE*	989.46	214 ± 1	-10.5
P	OVAL_CHICK	184-194	WEKAFKDEDTQ	1396.64	275 ± 2	-9.7
P/C	LYSC_CHICK	57-71	NTQATNRNTDGSTDY	1657.7	306 ± 3	-10.4
C	ALDOA_RABIT	44-57	QSIGTENTEENRRF*	1680.79	320 ± 1	-7.1
P	LYSC_CHICK	36-52	DNYRGYSLGNWVCAAKF	1963.91	351 ± 2	-8.1

Note: the table list the enzyme used for proteolysis (Enz., T = trypsin, C = chymotrypsin, P = pepsin, the UniProtKB protein identifier, position of peptide in the protein, amino acid sequence, monoisotopic mass-to-charge ratio (m/z), collision cross-section (Ω), and percent deviation from the average mobility-mass correlation (% Dev.). Collision cross-sections and reduced mobilities are reported as the average ± 1σ for 5 replicate measurements. (*) Indicates peptide sequences were confirmed by tandem MS.

Thus, it appears that the number of polar residues can be related to the conformations of peptide ions. That is, the presence of amino acids that can form a salt-bridge and/or H-bond interactions increases the packing efficiency of the peptide ions and reduces the collision cross-section. Attributing compact conformations to peptide ions that contain hydrophilic side-chains is indirectly supported by data for peptide ions that contain non-polar residues. The presence of non-polar amino acids appears to contribute more significantly to collision cross-sections than do polar groups, because non-polar side chains do not contribute to charge solvation or participate in charge-dipole or dipole-dipole interactions [78], or cation-π interactions.

The effect on helix formation of peptide ions in the gas phase

As mentioned previously, the peptide ions with secondary structures can be easily identified by their signals that positively deviated from the average drift time-m/z trend for peptides [56]. In the database we collected, 37 of 607 peptide ions have collision cross-sections that are larger than 7% from that expected for globular ions. These peptides corresponding to extended conformation are good candidates for helical structures. Jarrold and coworkers have reported the use of designed peptides to study helix formation of linear peptides in the gas-phase by IM-MS, and concluded three major factors influencing helix formation in un-solvated peptides [53]: i) the location of the residues with high helix propensity (*i.e.,* Val, Leu, Ala); ii) the number of neighboring residues with high helix propensity; iii) the location of the charge site. A positive charge at the C-term stabilizes the helix macro-dipole.

Different from Jarrold's studies which used the designed linear peptides, i.e., poly alanine, as model peptides, my study on helical structures of peptides in gas-phase focuses on the proteolytic digest fragments of proteins. Extensive molecular simulations have been used to examine helix formation in the twenty peptides with large collision cross-sections listed in Table 4. The results from simulated annealing studies are shown in Figures 2 and 3, which indicate that some of the peptides exhibit partial helical structures as gas-phase ions. The peptides with helix show similarity on their peptide composition: i) they have charge-carry residue at C-term. A helix has a macrodipole that results from the alignment of the backbone NH and CO groups. The positive end of the macrodipole is at the N terminus, and placing a positive charge at C terminus stabilizes the helix; ii) the peptide have at least one residue with high helix propensity in the middle part of the sequence. Val, Leu, Ala are the residues with high helix propensity. If they locate in the center of the peptide sequence, it can propagate helix formation to both ends. If they are adjacent to each other, it promotes helix formation as well. These observations are in a good agreement with Jarrold's conclusions [53].
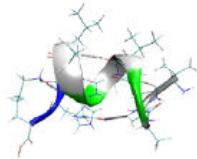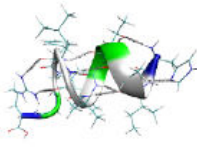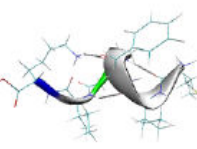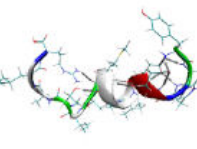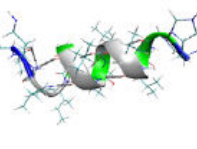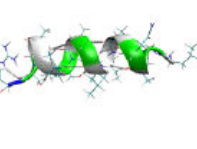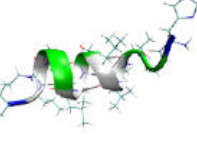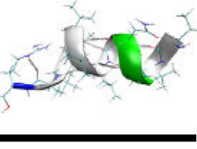
Sequence	Source	m/z	Ω exp (Å²)	Variance	Theoretical structure
VVAGVANALAHK	HBB_HORSE	1149.67	292 ± 2	9.8%	
HIATNAVLFFGR*	OVAL_CHICK	1345.74	327 ± 3	10.1%	
MIFAGIK*	CYC_HORSE	779.45	223 ± 4	10.2%	
YSHEEIAMATVTALR*	ALDOA_RABIT	1691.84	400 ± 1	15.5%	
HGTVVLTALGGILKK*	MYG_HORSE	1506.94	378 ± 5	17.8%	
LGANAILGVSLAASR*	ENO1_YEAST	1412.82	366 ± 2	19.2%	
HGTVVLTALGGILK*	MYG_HORSE	1378.84	361 ± 2	19.6%	
LLGNVLVVVLAR*	HBB_HORSE	1265.83	343 ± 2	20.6%	

Figure 2. Compilation of peptide ions with helical conformations identified by simulation annealing. Color in sequence denotes the secondary structure element. Green is turn, pink is helix, and black is random coil. Sequence with * indicates peptides were confirmed by tandem MS.
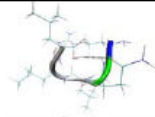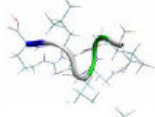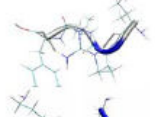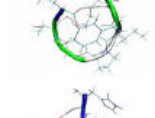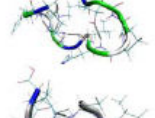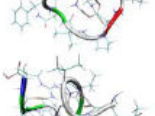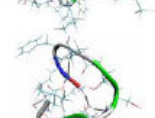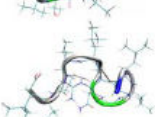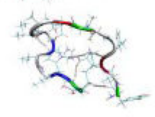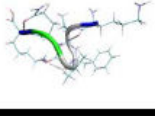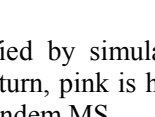
Sequence	Source	m/z	Ω exp (Å^2)	Variance	Theoretical structure
RQLLL	ALDOA_RABIT	642.43	192 ± 2	8.5%	
LNQLLR	ENO1_YEAST	756.47	215 ± 2	8.5%	
IRLF	MYG_HORSE	548.36	173 ± 1	8.6%	
KYIPGTK	CYC_HORSE	806.48	225 ± 4	8.6%	
HKTGPNLHGLFGR*	CYC_HORSE	1433.78	338 ± 3	8.9%	
ALELFR*	MYG_HORSE	748.44	215 ± 2	9.3%	
LRVDPVNFK	HBA_HORSE	1087.63	280 ± 2	9.4%	
LLVVYPWTQR*	HBB_HORSE	1274.73	315 ± 2	10.0%	
DKFLSSVSTVL	HBA_HORSE	1195.66	304 ± 2	11.2%	
RLLGNVLVVVL	HBB_HORSE	1194.79	304 ± 4	11.3%	
YLEFISDAIIHVLHSK*	MYG_HORSE	1885.02	419 ± 4	12.7%	
KIFVQK	CYC_HORSE	762.49	225 ± 3	12.9%	

Figure 3. Compilation of peptide ions with non-helical conformations identified by simulation annealing. Color in sequence denotes the secondary structure element. Green is turn, pink is helix, and black is random coil. Sequence with * indicates peptides were confirmed by tandem MS.

Compared to the peptides with helix, some of the non-helical peptides have the positive charge at the N terminus, which destabilizes the helix. For example, addition of HK to the TGPNLHGLFGR peptide, residues 28-38 of cytochrome c, results in dramatic changes in ion structure. The helical region of TGPNLHGLFGR spans the residues GPNLHGL, whereas the residues TGPNL of HKTGPNLHGLFGR, which correspond to residues of 26-38 or cytochrome c, adopt a turn structure (shown in Figure 4). The preference for helical vs. turn conformation can be a result from the location of charge, *i.e.,* the principle charge site for TGPNLHGLFGR is the C-terminal arginine, whereas the preferred charge carrying region of HKTGPNLHGLFGR is potentially the N-terminal HK residues [79]. This explanation is also supported by marked differences in the fragmentation spectra of the two peptides, *i.e.,* TGPNLHGLFGR $[M + H]^+$ ions dissociate to yield a much higher abundance of y_i type ions, owing to the C-terminal charge carrier, and HKTGPNLHGLFGR $[M + H]^+$ ions yield almost exclusively a_i and b_i type fragment ions. Such observations underscore the effects of peptide composition, especially charge-carrying residues, on the structure(s) of gas-phase peptide ions.

To further investigate the side chain contribution to helix formation in gas phase peptide ions, the short peptide with the defined helical conformation as gas-phase ion from the previously described database, *i.e.*, MIFAGIK, residues 80-86 of cytochrome c, has been used. The simulated annealing result suggests that MIFAGIK exhibit partial helix in the region of the residues IFA. The original charge site is blocked by different functional groups to verify the effects of charge location on the peptide structure in the gas phase. The detail will be presented in next chapter.

A)

B)
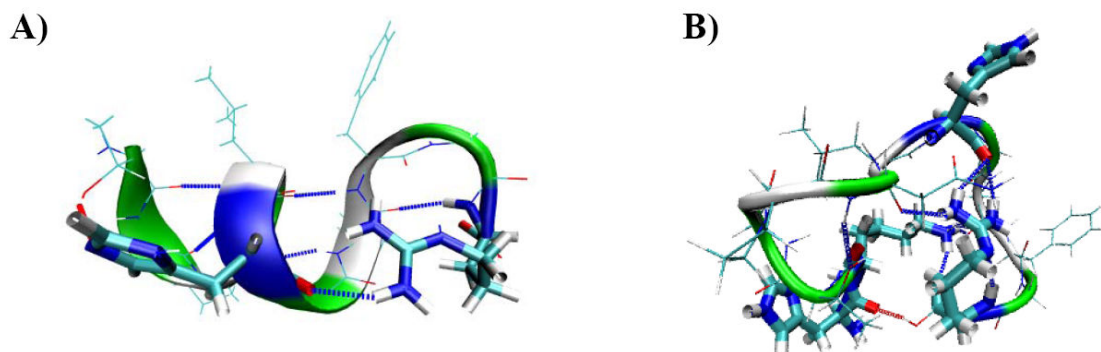
Figure 4. Proposed structures for TGPNLHGLFGR [M + H]⁺ ions (A) and HKTGPNLHGLFGR [M + H]⁺ ions (B) consistent with the ion-neutral collision cross-sections measured by MALDI. The images are produced using VMD (UIUC, Urbana-Champaign, IL). The colors represent different types of residues; white denotes non-polar residues, green denotes polar residues, blue denotes basic residues, and red denotes acidic residues. α–helical regions are given as coiled ribbons.

ESI vs. MALDI

Electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) are the most popular ion sources used for peptide and protein mass analysis, they are both soft ionization techniques by which the ions can be transferred intact into the gas phase. However, the ions formed by these techniques can be quite different in terms of internal energy and structure. Here we compare the experimental collision cross sections of peptide ions formed by different ionization processes and use computational simulations to probe the structural difference of peptide ions generated by ESI and MALDI.

Clemmer and coworkers have reported a collision cross-section database of 420 singly charged peptides generated by ESI-IM-TOFMS [51], which is quite consistent with the average globular mobility-mass correlation derived from the database we collected. The peptide ions identified by both ionization methods are listed in Table 6. For those peptides with same sequence, the collision cross-section acquired by MALDI is normally 3~7% larger than the one acquired by ESI. In a few cases, the difference is even larger than 10%. This can be explained by the different gas-phase conformations generated by MALDI and ESI. As a specific example, Clemmer reported a collision cross-section for TGPNLHGLFGR [M + H]$^+$ ions of 258 Å^2, but we obtained a value of 286 Å^2, a difference of ~ 9%. We performed extensive analysis of the data and are convinced that the differences are outside the bounds of experimental error. Using molecular modeling and molecular orbital calculations, we identified two helical structures (Figure 5) that agree very well with the experimental collision cross-sections.

Table 6. Compilation of peptide ions with the ion-neutral collision cross-section values measured by MALDI and ESI. Highlighted peptides are chosen for further simulation study. Peptide with * indicates the sequence was confirmed by tandem MS.

# aa	Protein	Sequence	m/z	$\Omega_{MALDI}(\text{Å}^2)$	$\Omega_{ESI}(\text{Å}^2)$	$\Delta\Omega$
4	ALBU_BOVIN	HKPK	509.32	145	150	-3.6%
6	ALDOA_RABIT	IVAPGK	584.38	167	173	-3.6%
4	ENO1_YEAST	VYAR	508.29	153	157	-2.9%
4	ALBU_BOVIN	FGER	508.25	145	148	-1.9%
4	CASB_BOVIN	KIEK	517.33	155	155	-0.3%
5	ALBU_BOVIN	IETMR	649.33	181	181	-0.3%
6	ENO1_YEAST	TGAPAR	572.32	163	163	0.1%
5	ENO1_YEAST	TGQIK	546.32	158	158	0.1%
6	ENO1_YEAST	AAAAEK	560.30	161	160	0.3%
7	ALDOA_RABIT	GGVVGIK	629.40	177	176	0.9%
6	ALBU_BOVIN	AWSVAR	689.37	187	185	1.1%
8	ALBU_BOVIN	AEFVEVTK	922.49	228	223	1.9%
7	ENO1_YEAST	GVLHAVK	723.45	203	199	2.3%
6	MYG_HORSE	NDIAAK	631.34	178	174	2.5%
7	CASB_BOVIN	VLPVPQK	780.50	212	207	2.5%
6	ALBU_BOVIN	SEIAHR	712.37	186	182	2.5%
6	MYG_HORSE	ASEDLK	662.34	180	175	2.8%
6	CASB_BOVIN	EMPFPK	748.37	199	193	2.9%
7	ENO1_YEAST	IATAIEK	745.45	209	203	3.1%
6	CASB_BOVIN	EAMAPK	646.32	182	176	3.4%
8	ALDOA_RABIT	RCQYVTEK	1026.50	253	245	3.6%
7	CASB_BOVIN	GPFPIIV	742.45	203	196	3.6%
6	ENO1_YEAST	SVYDSR	726.34	192	184	4.0%
7	ALDOA_RABIT	DGADFAK	723.33	193	186	4.1%
6	ENO1_YEAST	LNQLLR	756.47	214	205	4.5%
8	ALDOA_RABIT	ELSDIAHR*	940.48	232	221	4.6%
7	ALBU_BOVIN	VLASSAR	703.41	204	195	4.7%
7	ENO1_YEAST	HLADLSK	783.44	211	202	4.8%
9	ALDOA_RABIT	QLLLTADDR	1044.57	259	246	5.3%
10	ALBU_BOVIN	LVNELTEFAK	1163.63	282	268	5.5%
6	ENO1_YEAST	ANIDVK	659.37	187	177	5.5%
7	ENO1_YEAST	TFAEALR	807.44	222	210	5.7%
7	ALBU_BOVIN	YLYEIAR	927.49	242	228	6.2%
6	MYG_HORSE	ALELFR*	748.44	216	203	6.5%
6	MYG_HORSE	TEAEMK	708.32	194	183	6.5%
11	ALBU_BOVIN	HLVDEPQNLIK*	1305.72	302	283	6.7%
7	CASB_BOVIN	AVPYPQR	830.45	217	203	6.9%
7	CYC_HORSE	MIFAGIK*	779.45	222	207	7.1%
11	MYG_HORSE	LFTGHPETLEK*	1271.66	297	277	7.4%
8	MYG_HORSE	YKELGFQG*	941.47	245	228	7.5%
13	ALBU_BOVIN	LGEYGFQNALIVR*	1479.80	329	306	7.6%
14	CYC_HORSE	TGQAPGFTYTDANK*	1470.69	314	287	9.5%
15	ENO1_YEAST	LGANAILGVSLAASR*	1412.82	365	333	9.7%
11	CYC_HORSE	TGPNLHGLFGR*	1168.62	283	258	9.7%
12	HBA_HORSE	VVAGVANALAHK	1149.67	294	268	9.9%
9	HBA_HORSE	LHVDPENFR	1126.56	273	247	10.9%
10	HBA_HORSE	LLVVYPWTQR*	1274.73	317	283	12.1%
11	ALBU_BOVIN	HPEYAVSVLLR	1283.71	315	281	12.1%
9	HBA_HORSE	LRVDPVNFK	1087.63	281	251	12.2%
7	ENO1_YEAST	AAGHDGK	655.32	198	170	16.2%

A)

B)
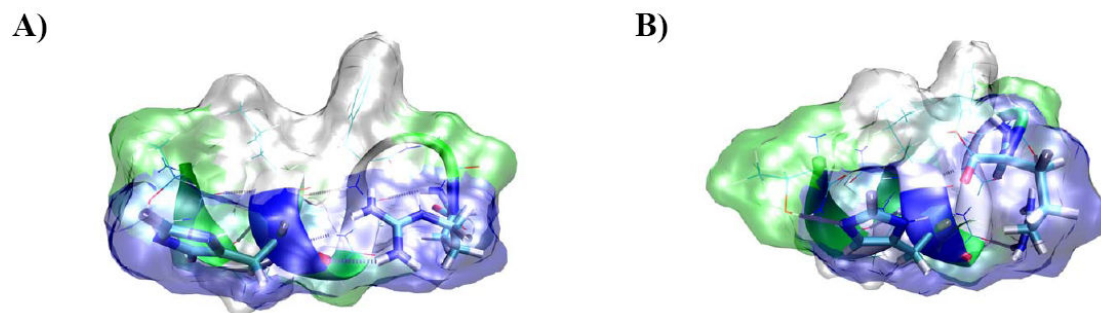
Figure 5. Proposed structures for TGPNLHGLFGR $[M + H]^{+}$ ions consistent with the ion-neutral collision cross-sections measured by (A) MALDI and reported for (B) ESI [18], respectively. The images are produced using VMD (UIUC, Urbana-Champaign, IL). The colors represent different types of residues; white denotes non-polar residues, green denotes polar residues, blue denotes basic residues, and red denotes acidic residues. α–helical regions are given as coiled ribbons.

Both structures appear to possess significant helical character, but the two structures differ in terms of the orientation of the side chains near the C-terminus. That is, the proposed structure of the MALDI formed ion contains a turn at the C-term owing to hydrogen bonding interactions between the guanidinium ion on the arginine side chain and glycine (G^{10}), which elongates the structure. This type of interaction is absent in the proposed ESI structure.

Extensive molecular modeling studies have been also implemented on other seven peptide ions where similar differences have been observed. These peptides are randomly picked from Table 6 with different sequence length. For each peptide ion, 3,600 candidate structures are generated from MD simulation, and the one that matches the experimental cross section value with the lowest energy is selected as the proposed structure for that particular peptide. The proposed structures for those peptide ions generated by MALDI and ESI are presented in Figure 6. An interesting observation is that, for short peptides with up to ~ 7 residues, the proposed structures of peptide ions formed by ESI and MALDI have pretty similar backbone structures and their large difference in collision cross sections can be a result from the different projection of side chains in space. However, for longer peptides, since they have more degrees of freedom for positioning structural elements in backbone; both side chain orientation and backbone structure attribute to the difference in collision cross sections between ESI and MALDI.
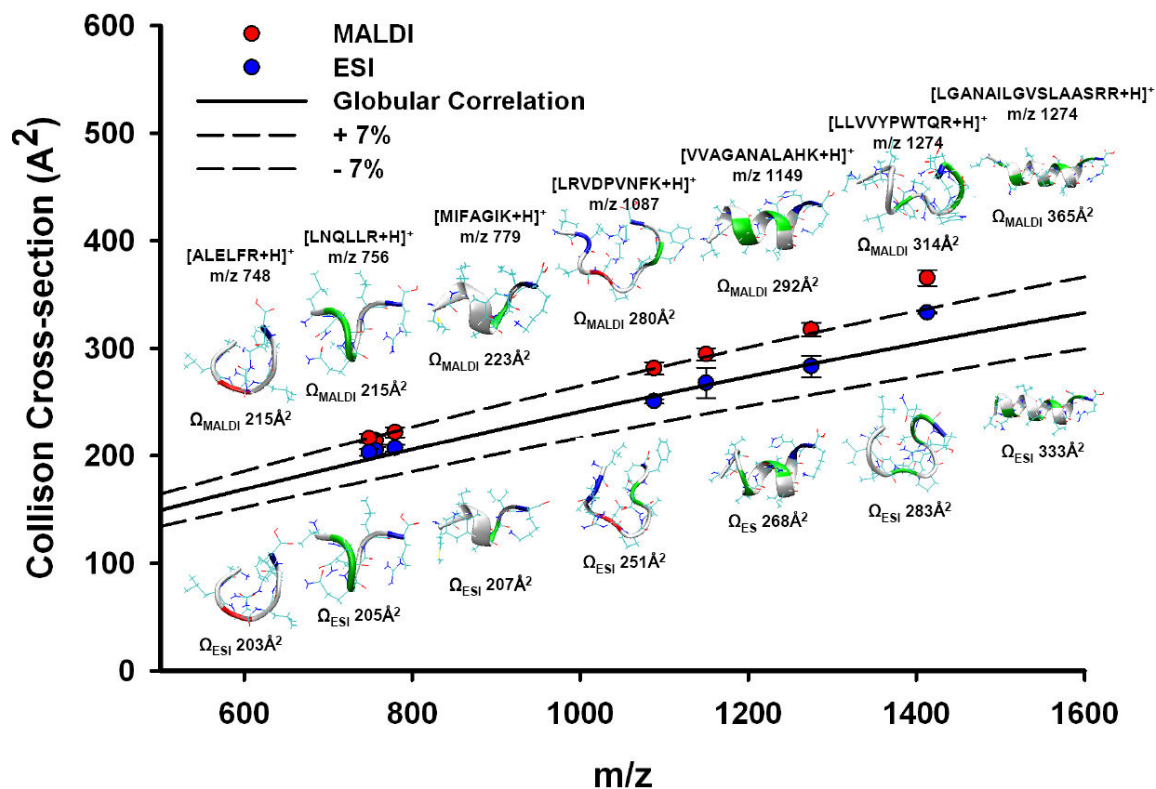
Figure 6. Proposed structures for 7 peptide $[M + H]^+$ ions consistent with the ion-neutral collision cross-sections measured by ESI and MALDI. The images are produced using VMD (UIUC, Urbana-Champaign, IL). The colors represent different types of residues; white denotes non-polar residues, green denotes polar residues, blue denotes basic residues, and red denotes acidic residues. α–helical regions are given as coiled ribbons.

Conclusion

Ion-neutral collision cross-sections for 607 $[M + H]^+$ peptide ions are compiled into a dataset containing structure and mass information using MALDI-IM-TOFMS. The peptides included in the database were generated by enzymatic digestion of known proteins using three different enzymes, resulting in peptides that differ in terms of amino acid composition as well as N-terminal and C-terminal residues. The majority (63%) of the peptide ion collision cross-sections correlate well with structures that are best described as charge-solvated globules, but a significant number of the peptide ions exhibit collision cross-sections that are significantly larger or smaller than the average, globular mobility-mass correlation. Of the peptide ions having larger than average collision cross-sections, approximately 71% are derived from trypsin digestion (C-terminal Arg or Lys residues) and most of peptide ions that have smaller (than globular) collision cross-sections are derived from pepsin digestion (90%). Molecular simulations results suggest that peptide ions having larger than predicted collision cross-sections correspond to extended conformations and can be good candidates for helical structures. Whereas the peptides with smaller cross-sections would prefer more compact structure due to the salvation interaction from their polar residues. Compared to the peptide collision cross-sections measured using electrospray, we find reasonably good correlation between the measured values, but in some cases the collision cross-sections differ by as much as 3-10%. Simulation results indicate that ESI yields gas-phase ions which can differ in terms of structure or conformation from those formed by MALDI. Since the ESI and MALDI IM-MS data were acquired under different operation

parameters using different instruments, the question concerning whether the different results are related to the nature of the sample, the specific details of the sample preparation, or the instrumental parameters used in operation remains unanswered. To avoid the errors generated by different operation systems, the mobility measurements for peptide ions generated by ESI and MALDI should be performed using same operating condition on same mobility instrument. Now the mobility instrument with both ionization sources is under development in our lab.

CHAPTER III

CONTRIBUTIONS OF MOLECULAR FRAMEWORK TO IMS

COLLISION CROSS-SECTIONS OF GAS-PHASE PEPTIDE IONS*

Introduction

The emphasis of mass spectrometry based biological chemistry is shifting from compound identification to structural studies of large biomolecules and biomolecule complexes [9-15], including membrane proteins [16]. The next phase of 'omics' related research must be aimed at obtaining and predicting additional dimensions of information, such as secondary, tertiary, and quaternary structures and linkage specific information for glycans. Although sophisticated structural characterization tools such as NMR and XRD provide the most information, high throughput analysis of complex biological mixtures obtained by using these techniques is an underdeveloped technology. On the other hand, IMS is much more than a separation device, the structural information derived from 2D conformation space afforded by IM-MS is potentially well-suited to both high throughput applications and complex biological samples.

A number of laboratories have focused their research on developing IM-MS for biophysical studies of peptides and proteins [34,52,55,56,63,64]. In previous work we showed that a large proportion of singly charged peptide ions (formed by MALDI) appear on a single trendline in mobility-m/z plots, *i.e.,* plots of arrival-time distribution

_____

*Reprinted with permission from "The Contributions of Molecular Framework to IMS Collision Cross-Sections of Gas-phase Peptide Ions" by Tao, L.; Dahl, D. B.; Perez, L. M.; Russell, D. H., 2009, *Journal of the American Society for Mass Spectrometry*, 20, 1593-1602, Copyright [2009] by Elsevier.

(ATD) or ion-neutral collision cross-section (Ω) vs. m/z [54];however, a small number of ion signals deviate (>3 to ~20%) from the expected trendline, and non-peptidic ion signals appear on separate, compound class specific trendlines [34, 54]. Ruotolo *et al.* showed that gas-phase $[M + H]^+$ ions of LLGNVLVVVLAR (derived from bovine hemoglobin) prefer extended (helical) structure(s) resulting in a larger collision cross-section than random coil structures having the same or similar m/z values [55, 56], while some post-translationally modified (PTM) peptide ions (phosphopeptides) tend to pack more tightly than the unmodified protonated peptide ions owing to intra-molecular charge-solvation and/or formation of salt-bridged type structures [57, 58]. In addition, we have used chemical derivatization via acetylation of the N-terminus and internal basic lysine residues and methylation of the acidic glutamic residues to show that the helical propensity of a given peptide can be increased by reducing the number of 'salt-bridge' intramolecular interactions [59].

Molecular dynamics (MD) is the method of choice for correlating the measured collision cross-section (Ω_{meas}) with candidate structures. Ion structure is derived by comparing Ω_{meas} with the calculated collision cross-section (Ω_{calc}) of the lowest energy structure obtained using MD simulations. This approach involves selecting the 'lowest energy structure' from a large pool of candidate structures generated through multiple tiers of simulated annealing [80]. It is important to note, however, the lowest energy structure may not be representative of Ω_{meas} since under the experimental conditions, *i.e.*, long resident times in the drift cell and small amounts of collisional heating can facilitate rearrangement reactions that may be subject to low energy barriers. The ion population

is composed of an ensemble of many conformations at a defined ion effective temperature range as opposed to a single conformation assumed at the lowest energy [81].

The challenge for structure IM-MS is to design gas-phase experiments that critically evaluate the structural assignments [60-62]. The research described here is aimed at bridging the gap between the experimental domain of IM-MS and the MD simulations and bioinformatic tools that are essential to interpretation of the data. In this study, we apply a novel clustering algorithm to a model peptide to identify groups of structural elements from a large pool of diverse candidate structures. The clustering algorithm is similar to that used by Damsbo *et al*. [82], *i.e.*, have grouped the candidate structures based on the similarity of backbone structure; however, our procedure provides estimates of the uncertainty of the cluster membership and the degree of purity of the cluster.

In this study, we use NH_2-Met-Ile-Phe-Ala-Gly-Ile-Lys-COOH (MIFAGIK), residues 80-86 of cytochrome *c*, as our model peptide to develop methods to evaluate the complete candidate structure population. We selected this peptide because the Ω_{meas} for $[M + H]^+$ ions formed by MALDI is about 5% larger than that expected for globular conformation [54] and smaller than that expected for helical structure. The observation raises the question of whether the ion population is composed of a number of very similar conformations and can we use statistical analysis tools to gain information about structural diversity. Although simulated annealing experiments yield a lowest energy structure that contains a helical turn conformation in the region of the residues Ile^2-Phe^3-

Ala^4, candidate structures with Ω_{calc} that fall within ±2% of the measured value are comprised of a number of structure types.   We have also compared MIFAGIK derivatives, including acetylated, Boc and Fmoc protected N-terminus, C-terminal methyl ester, Fmoc protected Lys^7 and $[M + Na]^+$ ions. The experimental data provides new insight into the role of protecting groups in the N- and C-terminal positions and the role of the charge site in the conformational preference of small peptides.

Experimental

Materials

The peptide MIFAGIK (MW 778.5) was purchased from Genscript Corp. (Piscataway, NJ) and used without further purification. Bradykinin (RPPGFSPF, MW 1060.2), substance P (RPKPQQFFGLM-NH$_2$, MW 1346.6), anhydrous methanol, acetyl chloride, acetic anhydride boc anhydride, 9-fluorenylmethyl chloroformate and dioxane were obtained from Sigma (St Louis, MO). The peptide derivatives were synthesized using conventional solution phase methods [83, 84]. The acetyl, Boc and Fmoc group were used for amino group protection and the C-terminus was protected by the methyl ester. Deprotection of Boc group was performed using hydrochloric acid. All peptide sequences were confirmed using tandem mass spectrometry (Applied Biosystems 4700 Proteomics analyzer).

MIFAGIK-OMe

The methanolic HCl reagent was prepared by dropwise addition of 800 µL acetyl chloride to 5 mL of anhydrous methanol with stirring. After 5 min, 200 µL methanolic HCl was added to 1 mg of dry peptide. The solution was incubated for a period of 2 h at room temperature and was dried by lyophilization. The resulting O-methyl ester was dissolved in 1 mL distilled water and used for mass analysis without further purification.

N-Ac-MIFAGIK

The stock solution of peptide MIFAGIK was prepared by dissolving 1 mg peptide in 1 mL of 50 mM ammonium bicarbonate. The acetylation reagent was prepared by mixing 10 µL acetyl anhydride with 20 µL anhydrous methanol and then added to 50 µL peptide stock solution. The reaction was allowed to proceed for a period of 15 min at room temperature. The sample was dried by lyophilization. The peptide solution was reconstituted by mixing with 50 µL distilled water and analyzed by mass spectrometry without further purification.

N-Fmoc-MIFAGIK

The Fmoc group was introduced by treatment of 10 µL free peptide MIFAGIK stock solution with 10 µg 9-fluorenylmethyl chloroformate in 10 µL aqueous dioxane. The reaction was allowed to proceed for a period of 60 min at room temperature. The sample was dried by lyophilization. The peptide solution was reconstituted by mixing

with 20 µL distilled water and analyzed by mass spectrometry without further purification.

MIFAGIK-N-Fmoc

The MIFAGIK-*N-Fmoc* derivative was synthesized by first protecting the N-terminus with a Boc protecting group and then reacting with Fmoc followed by removable of the Boc group under acidic conditions. A solution of Boc-anhydride (2 µmoL) in 10 µL anhydrous methanol was added to 10 µL free peptide MIFAGIK stock solution. After stirring for 30 min at room temperature, the mixture was concentrated in vacuo. The residue was dissolved in 10 µL aqueous dioxane in the presence of sodium carbonate to which 10 µg 9-fluorenylmethyl chloroformate in 10 µL aqueous dioxane was added. The reaction was allowed to proceed for 60 min at room temperature. The Boc group was removed by addition of 5 µL of 1.2 M HCl, which was allowed to react for 10 minutes. The product was then analyzed by mass spectrometry without further purification.

MALDI-IM-TOFMS

Stock solutions of peptide (1mg/mL) were mixed with re-crystallized α-cyano-4-hydroxycinnamic acid (Sigma) in a molar ratio of ~2000:1, and 1 µl aliquots of the peptide/matrix mixture was deposited on a stainless steel MALDI sample plate.

The MALDI-IM-TOFMS analysis was performed using a home-built instrument previously described [68]. Briefly, MALDI was performed using a frequency tripled

(355 nm) Nd:YAG laser (CrystaLaser, Reno, NV) operated at a pulse rate of 200-400 Hz. The resulting ions were introduced into a 15 cm drift tube maintained at a pressure of ca. 2.7 Torr of He maintained at room temperature. Under these conditions the E/p ratios range from 18 to 34 $Vcm^{-1}torr^{-1}$, which provides mobility resolution of 15 to 25. The ions eluting the drift cell were extracted and mass analyzed by an orthogonal reflectron-TOF, typical m/z resolution ca. 2000 - 4000. The mass spectrometer was externally calibrated using two-point calibration on C_{60} (M_r = 720) and C_{70} (M_r = 840) radical cations (Sigma) [5]. The measurements of collision cross sections were externally calibrated with $[M + H]^+$ ions of bradykinin (Ω_{meas} = 245 $Å^2$) and substance P (Ω_{meas} = 292 $Å^2$) [35]. The 2D IM-MS data was acquired and processed by using custom software (Ionwerks, Inc.).

The experimental ion-neutral collision cross-sections were determined from the empirical drift times (t_d) by the following equation [69]:

$$\Omega_{meas} = \frac{(18\pi)^{\frac{1}{2}}}{16} \frac{ze}{(k_b T)^{\frac{1}{2}}} \left(\frac{1}{m_I} + \frac{1}{m_B} \right)^{\frac{1}{2}} \frac{t_d E}{L} \frac{760}{P} \frac{T}{273.2} \frac{1}{N_0} \tag{1}$$

where z is the charge of the ion, e is elementary charge, N_0 is the number density of the drift gas at STP, k_b is Boltzmann's constant, m_I is the mass of the ion, m_B is the mass of buffer gas, E is the IM electric field strength, L is the drift tube length, P is the buffer gas pressure and T is the system temperature.

Molecular dynamics

Molecular dynamics simulations were performed as described previously [56, 73]. Peptide structures were first constructed using Insight II 2000(Accelerys, San Diego, CA). Starting structures were limited to 4 different structures: an extended form, α-helix form, 3_{10}-helix form and Pi-helix form. For the peptide $[M + H]^+$ ions, we assumed that the proton was on the most basic residue, Lys and N-terminus, whereas for the $[M + Na]^+$ ions, Na^+ was placed in a central position on the peptide backbone. Simulated annealing was performed with the OFF program in Cerius 4.9 (Accelrys, San Diego, CA) using the Consistent Force Field (CFF 1.02). During the simulation, the starting structures were gradually heated over the course of 280 ps in a step-wise fashion, (relaxation time is 0.1 ps and time step is 0.001 ps), starting and ending at 300 K and peaking at 1000 K. After each annealing cycle, a minimized structure was produced. Annealing cycles were repeated 300 times generating 300 minimized structures for each starting structure. Eight more structures were selected from first stage results as the starting structures of the second stage simulations, creating a total of 3,600 minimized structures for each peptide. Ω_{calc} of each minimized structure was calculated using the elastic hard sphere scattering (EHSS) method in the MOBCAL software [74].

Dynamic clustering algorithm

We develop a novel clustering procedure as an integral part of our methodology to correlate IMS collision cross-section data with candidate structures. Our approach uses an ensemble of many conformations rather than simply the lowest energy structure.

Damsbo *et al.* [85] also use clustering methods for this purpose, although our approach differs from theirs in several respects. First, the pairwise distance for our clustering procedure is the root-mean-squared-deviation (RMSD) of the backbone coordinates calculated between all pairs of simulated structures that correlate to ±2% of Ω_{meas}, whereas Damsbo *et al.* [85] defined the pairwise distance in terms of backbone torsion angles. Even more fundamental, however, is the fact that the clustering procedures themselves are quite different. Damsbo *et al.* use the *k*-means clustering procedure which, depending on the initial starting conditions, may converge to sub-optimal configurations. Even if the optimal clustering is found, however, *k*-means only produces a single clustering estimate. Our novel clustering procedure, while more computationally intensive, is robust to the initial starting configuration. Further, rather than merely giving a single clustering estimate, our procedure provides a distribution of clustering. This allows us to investigate important aspects of the clustering distribution, including assessing the proximity of clusters and measures of clustering uncertainty.

Intuitively, our clustering procedure repeatedly reallocates structures among existing or new clusters based on the proximity (as measured by RMSD) of the structures to each other. For any given structure, the probability it is allocated to an existing cluster is a function of the RMSD values between this structure and those already in the existing cluster. Rather than being forced into the cluster with the highest probability, however, a structure is randomly allocated to clusters based on these probabilities. Further, there is also a chance that a structure may form a new cluster. A key aspect of the algorithm is that the clustering is dynamic in that structures are

reallocated among the existing and new clusters in each iteration. Repeated iterations provide a clustering distribution, from which a point estimate of clustering can be obtained and clustering uncertainly can be assessed. As we describe below, we use least-squares clustering [86] to obtain a point estimate and we assess uncertainty using the resulting pairwise probability matrix for whether two structures occupy the same cluster.

Formally, the dynamic clustering procedure can be explained as follows. A clustering of n structures can be represented by a set partition $\pi = \{S_1, \ldots, S_q\}$ of a set $S_0 = \{1,\ldots,n\}$, where the subsets are nonempty, mutually exclusive, and exhaustive. Two structures k and k' are clustered in cluster S if and only both k and k' are in S. The algorithm uses a Markov chain [86] starting at an initially clustering π and updating according to the following transition rule. For $k = 1, \ldots, n$,

$$p(k \in S \mid \cdot) \propto h_k(S), \text{ for all } S \in \pi$$

$$p(k \in S^* \mid \cdot) \propto \alpha,$$

where S^* is an new cluster not currently associated with any other structures and $h_k(S)$ is defined as follows. Let $d_{kk'}$ be the pairwise RMSD between structures k and k'. Let $w_{kk'} \propto (d^* - d_{kk'})^t$ be a measure of proximity of items k and k', where d^* is the overall maximum pairwise distance plus a small increment to ensure that all weights are strictly positive. Note that t has the effect of dampening or accentuating the proximity measurements. For each $k = 1, \ldots, n$, scale $w_{k1}, \ldots, w_{kn}$ such that $\sum_{k' \neq k} w_{kk'} = n - 1$. Finally, define $h_k(S) = \sum_{k' \in S} w_{kk'}$. Note that the probability of assigning a structure k to a

new cluster is $\alpha/(\alpha + n - 1)$ and the probably to being assigned to an existing cluster is $(n - 1)/(\alpha + n - 1)$.

A key theoretical consideration is whether this procedure has an equilibrium distribution. This Markov chain has a finite state space since the number of possible clustering, given by the Bell number, is finite. Since every clustering is accessible from every other clustering through repeated application of the transition rule, the Markov chain is irreducible. The Markov chain is aperiodic because, for every clustering, it is possible that an application of the transition rule will lead to the same clustering. Finally, since the weights are strictly positive, there is positive probability of returning to every clustering and the Markov chain is recurrent. Since this Markov chain is irreducible, recurrent, and aperiodic, it indeed satisfies the conditions to have an equilibrium distribution.

Following repeated application of the transition rule, many candidate clusterings have been generated. The pairwise probability $\hat{p}_{kk'}$ that two structures k and k' are clustered is estimated by the relative frequency among the candidate clusterings that k and k' occupy the same cluster. The least-squared clustering [86] selects the candidate clustering closest to the estimated pairwise probabilities in terms of squared distances:

$$\pi^{LS} = \underset{\pi \in \left\{ \pi^{(1)},...,\pi^{(B)} \right\}}{\operatorname{argmin}} \sum_{k=1}^{n} \sum_{k'=1}^{n} \left(\delta_{kk'}(\pi) - \hat{p}_{kk'} \right)^2$$

where $\delta_{kk'}(\pi) = 1$ if k and k' occupy the sample cluster in clustering π, and 0 otherwise. This method minimizes of a posterior expected loss of Binder [87] with equal costs of clustering mistakes.

To get an indication of clustering uncertainty, we propose to quantity we call the c-value. The c-value for a structure k in cluster S is the average pairwise probability of being clustered with the other structures in cluster S, i.e.,

$$c-value_k = \frac{\sum_{k' \in S} \hat{p}_{kk'}}{|S|},$$

where $|S|$ is the size the cluster S. If an item has a high c-value, there is more confidence that structure is clustered appropriately.

The clustering uncertainty can also be assessed by plotting a pairwise probability matrix. Arrange the rows and columns by the clustering as indicated by the least squares clustering and make the color of each element indicate the value of the estimated pairwise probability. This plot makes it easy to see what clusters are well defined and which clusters are closely related or very different from other clusters. We give an example of this plot later in the paper.

In implementing the procedure in this paper, we set $\alpha = 1$ and t at a value between 15 and 45 so as to provide a few well-separated clustering. We applied the transition rule about 500,000 times. In repeated application of the algorithm to the same dataset, the resulting clustering very similar as measured by the adjusted Rand index [88].

Results and Discussion

[M + H]$^+$ ions of MIFAGIK

All Ω_{meas} of the [M + H]$^+$ and [M + Na]$^+$ ions of MIFAGIK and the various derivatives are listed in Table 7. The [M + H]$^+$ ions of MIFAGIK ($\Omega_{meas} = 213 \pm 4$ Å^2) is

Table 7. Ion-neutral collision cross sections of peptide ions measured by MALDI-IM-TOFMS are determined from empirical drift times (t_d) using a hard sphere approximation. * Errors for all measured collision cross-section values are around ±2%.

Name	Squence	m/z	Ω_{meas}* (Å^2)
[M+H]$^+$	MIFAGIK	779.45	213
	MIFAGIK-O-Met	793.46	216
	Ac-N-MIFAGIK	821.48	223
	Boc-N-MIFAGIK	879.50	239
	Fmoc-N-MIFAGIK	1001.52	251
	MIFAGIK-Fmoc	1001.52	248
[M+Na]$^+$	MIFAGIK+Na	801.44	217
	MIFAGIK-O-Met+Na	815.45	223
	Ac-N-MIFAGIK-Ac+Na	885.47	237

approximately 5% larger than that expected for ions that fall on the globular mobility-mass trendlines [54], which suggests that the ion structures are somewhat elongated (partial helix). MD simulations as described above yield a total of 3,600 candidate structures for the $[M + H]^+$ ions and 631 fall within $\pm 2\%$ of Ω_{meas}. Figure 7a contains a scatter plot of energy versus Ω_{calc} generated by annealing molecular dynamics.

The dynamic clustering procedure described above was applied to the 631 structures that fall within $\pm 2\%$ of Ω_{meas}. The structural similarities among the candidate structures are revealed by plotting the pairwise probability matrix (Figure 7b), where the color of the element (i, j) indicates the estimated probability that structures i and j have similar backbone coordinates. It is possible that structures that fall within a given cluster may in fact share structural elements with another cluster. For example, cluster 1 contains conformations that are similar to each other as well as cluster 5, whereas the structures represented by cluster 2 are very different from those of cluster 5 and cluster 3. (shown in Figure 8). That is, the tan color indicates high probability of being clustered (i.e., having similar structures), whereas the off-diagonal blocks of dark colors (i.e., dark blue or purple) indicate the clusters do not share structural features. Eleven potential clusters were generated for the 631 candidate structures; however, clusters 1, 5, 2, and 3 contain over 85% of the total structural elements of the ion populations. Backbone projections of all candidate peptide structures within the most populated clusters are overlapped and displayed in Figure 8. The overlaid backbone structures of each cluster show random fluctuations, both in terms of backbone coordinates as well as positioning of the side chains. A representative structure for each cluster is obtained by selecting the

a)

b)
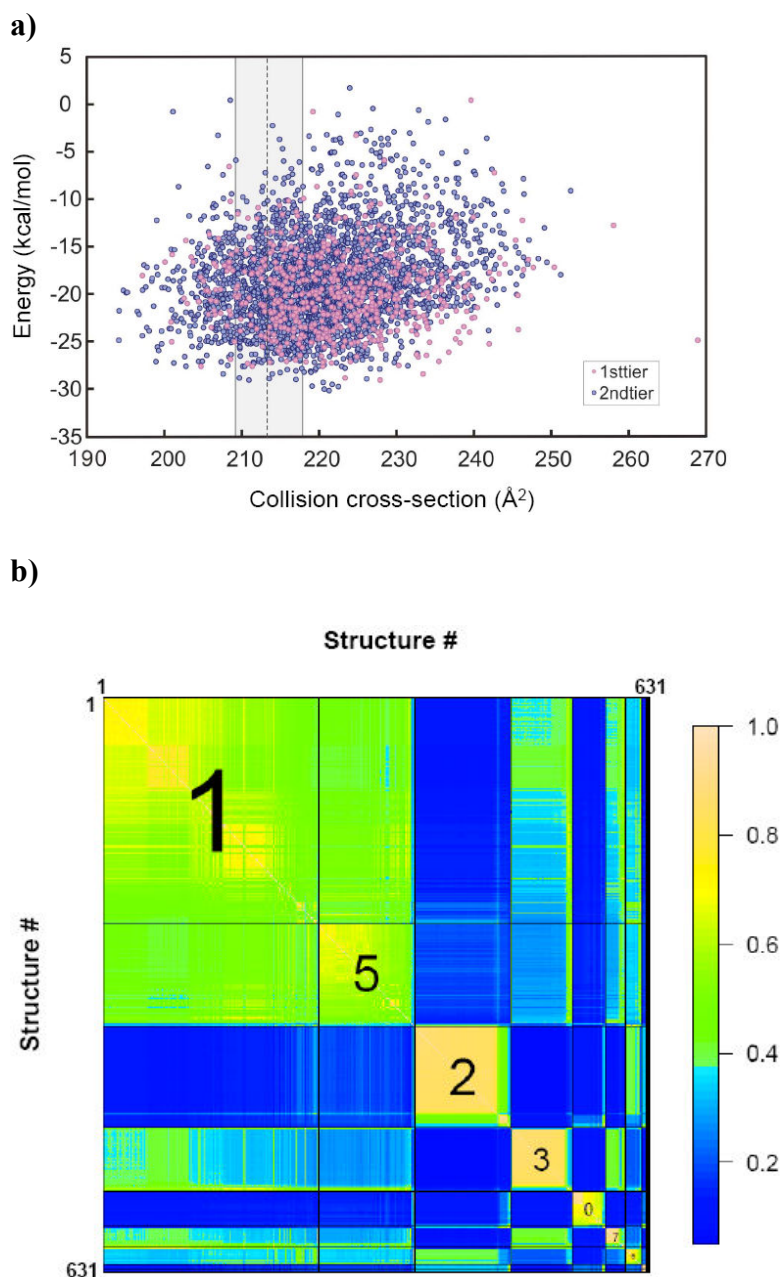
Figure 7. a) Scatter plot of MD energy versus calculated ion-neutral collision cross-section generated by annealing molecular dynamics for [MIFAGIK + H]$^+$ ions. The shaded area indicates ±2% deviation of the measured ion-neutral collision cross-sections, b) the pairwise probability plot of the MIFAGIK peptide ion.

Cluster	Backbone Structures	Population Percentage	Representative Structure	Structure Elements	High c-value Structures	C-values
1		39		α-turn (MIFAG)		0.616076 0.616042 0.615796
5		18		Random Coil		0.799788 0.799544 0.799392
2		18		β-turn (IFAG)		0.796034 0.795719 0.795341
3		11		α-helix (IFA)		0.627372 0.626751 0.62523

Population: 631 candidate structures

Figure 8. Backbone structures of four most populated clusters for $[MIFAGIK + H]^+$ ions classified by their RMSD values and the structure with highest c-value (representative structure) in each cluster.

structure with the highest c-value in the cluster. The c-value provides an indication of the certainty for each structure in the cluster. That is, the structure with a high c-value indicates a high confidence of being clustered appropriately; thus the structure with the highest c-value should be the best estimate of a cluster.

Note that clusters 1, 5, 2, and 3 represent different conformations owing to coordination of the charge site and specific backbone carboxyl groups. For example, cluster 1 is comprised of structures having a turn in the Met^1-Ile^2-Phe^3-Ala^4-Gly^5 sequence; cluster 5 is best described as a random coil; cluster 2 is characterized by a β-turn over Ile^2-Phe^3-Ala^4-Gly^5 sequence and cluster 3 corresponds to partial helical character over Ile^2-Phe^3-Ala^4 sequence. Note that 'the lowest energy structure' within the Ω_{meas} corresponds to a helical turn conformation over the IFA sequence, similar to cluster 3; however, this structure comprises a minor portion of the total ion population.

Our ultimate goal would be to evaluate the ion structural distribution as a function of energy; however, structure-energy distributions are a function of both energy and reaction dynamics and our current methodologies do not allow for such detailed information. Higher-level calculations to yield meaningful free energy values for pursuing this important question [81] will present in next chapter. Nonetheless, it is interesting to compare the range of energies for each conformation for the 4 most populated clusters. For example, ~80% of the structures in cluster 3 (α-helix; 11% of the ion population) have energies (Figure 9) that are lower than -20 kcal/mol, which suggests that a greater fraction of the low internal energy ions prefer cluster 3, whereas cluster 1 (α-turn; 39% of the ion population) is comprised of higher energy conformers.
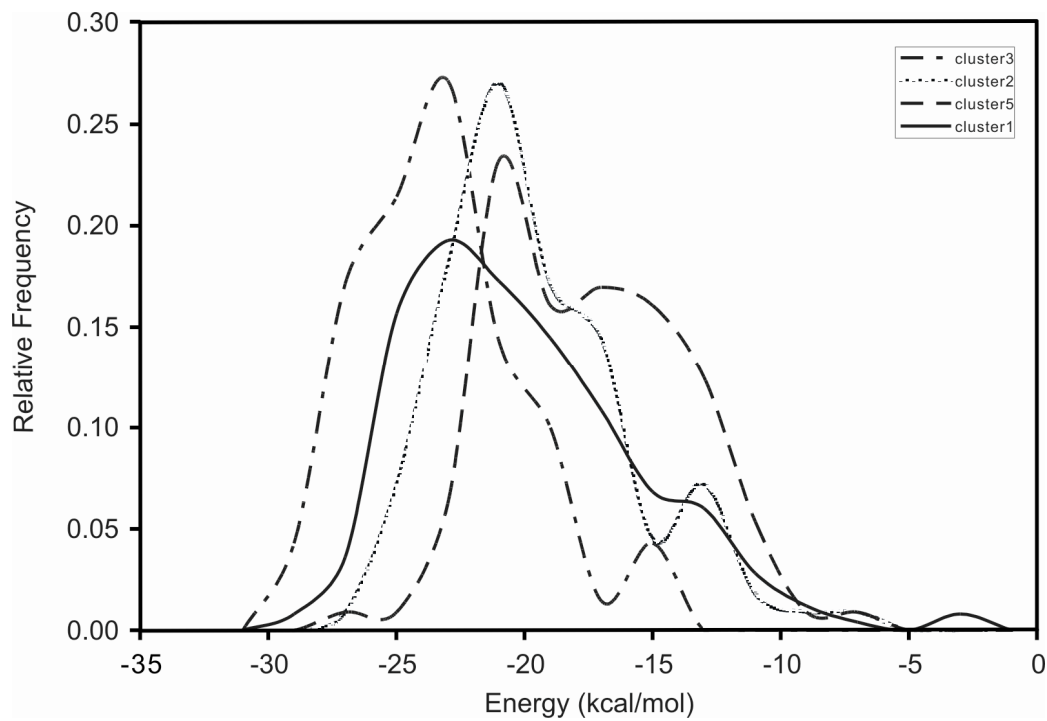
Figure 9. Energy profiles of four most populated clusters for $[MIFAGIK + H]^+$ ions.

These data raise an obvious question, is the structural population among the various clusters dependent on ion internal energy? Such questions can be probed by producing a greater population of low internal ions or possibly by performing variable-temperature IMS experiments. At the present time we do not have experimental capabilities for either experiment. Clemmer *et al*. have measured collision cross-sections for MIFAGIK $[M + H]^+$ ions formed by ESI-IMMS, and the Ω_{meas} of 207 ± 4 Å^2 [51] differs by 4% from that measured by MALDI. A total of 458 candidate structures that fall within ±2% of the ESI Ω_{meas} are subjected to cluster analysis (Figure 10). Note that the most populated clusters are quite similar in terms of backbone conformations to those formed by MALDI, *i.e*., approximately 34% of the structures have a turn in the Met^1-Ile^2-Phe^3-Ala^4-Gly^5 sequence. The β-turn structure spanning Ile^2-Phe^3-Ala^4-Gly^5 and FAGI sequence are also favored. The partial helical structure over Ile^2-Phe^3-Ala^4 sequence is the least abundant. Thus it appears that that the structures of peptide ions formed by ESI and MALDI have similar backbone structures and the difference in collision cross-sections are most likely a result of different projections of side chains; however, it is equally likely that these differences are simply variations in the experimental results. We are currently developing IM-MS instruments that will allow us to perform more extensive comparisons of MALDI and ESI formed ions as well as VT(~90-650K)-IM-MS experiments.

a)

b)

Population: 458 candidate structures
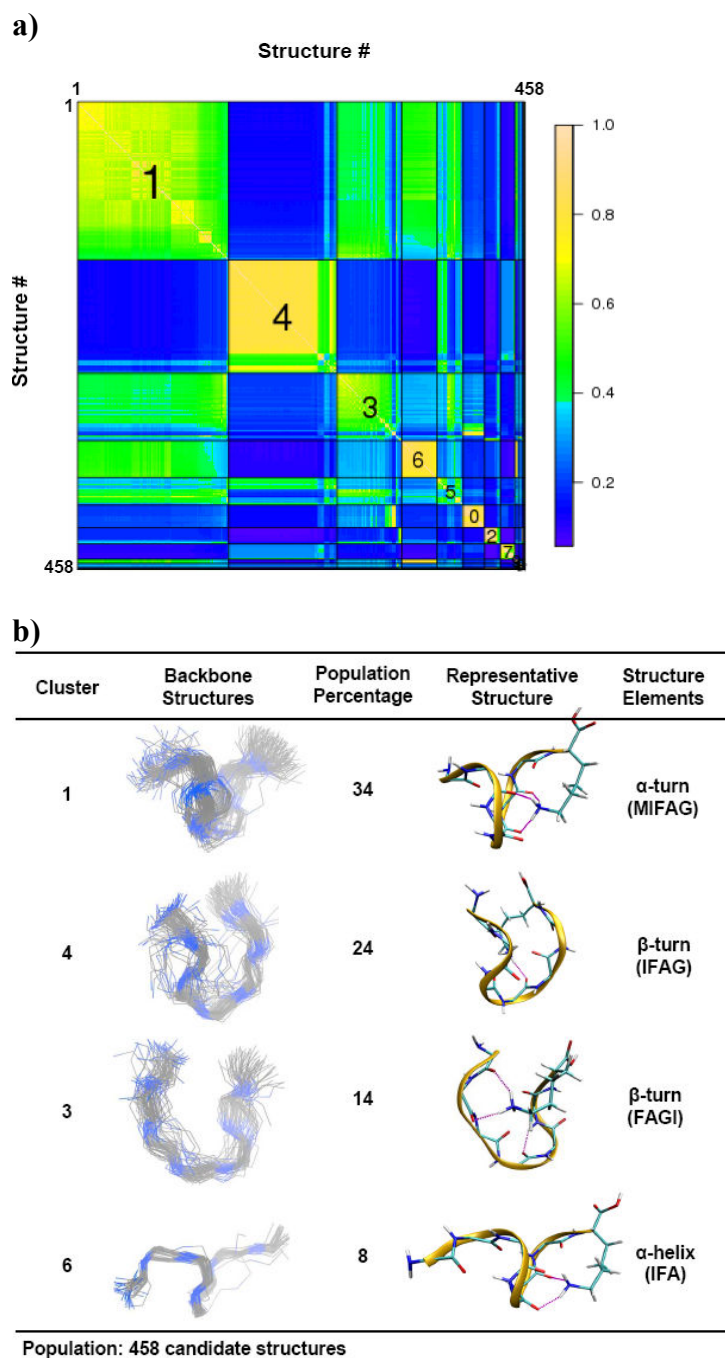
Figure 10. a) The pairwise probability plot of candidate structures for $[MIFAGIK + H]^+$ ions formed by ESI-IMMS[*], and b) the backbone structures of four most populated clusters for $[MIFAGIK + H]^+$ ions formed by ESI-IMMS[*] . [*] Ω_{meas} of $[MIFAGIK + H]^+$ was taken from data published by Clemmer *et al.* [35].

[M + Na]⁺ ion of MIFAGIK

The charge site can also affect the conformations of gas-phase peptide ions. Previous studies have shown that the dissociation reactions of $[M + H]^+$ and $[M + Na]^+$ ions differ significantly owing to different charge locations, H^+ prefer basic sites and Na^+ is very oxyphilic [89]. The MD simulations of MIFAGIK $[M + Na]^+$ ions were performed by placing Na^+ ion on the neutral peptide backbone, and the resulting energy vs. collision cross-section plot was similar to that shown in Figure 8. The candidate structures that fall within ±2% of the Ω_{meas} (217 ± 4 Å^2) (a total of 654 conformations) are subjected to cluster analysis. Cluster analysis yields a total four clusters, and two of these clusters contain approximately 84% of the total ensemble of representative structures (see Figure 11). These structures are characteristic of random coil amide backbones with the Na^+ ion surrounded by the amide groups in what appears to be a 'charge-solvated' structure. When superimposed the two conformations comprising cluster 1 and 3 appear to be mirror images. Cluster analysis is performed on $[N\text{-}Ac\text{-}MIFAGIK\text{-}Ac + Na]^+$ and $[MIFAGIK\text{-}O\text{-}Me + Na]^+$ ions as well, the Ω_{meas} are listed in Table 7 and cluster analysis results are shown in Figures 12 and 13. These ions prefer random coil conformation that are similar to those for $[MIFAGIK + Na]^+$ ions.

a)

b)

Population: 654 candidate structures
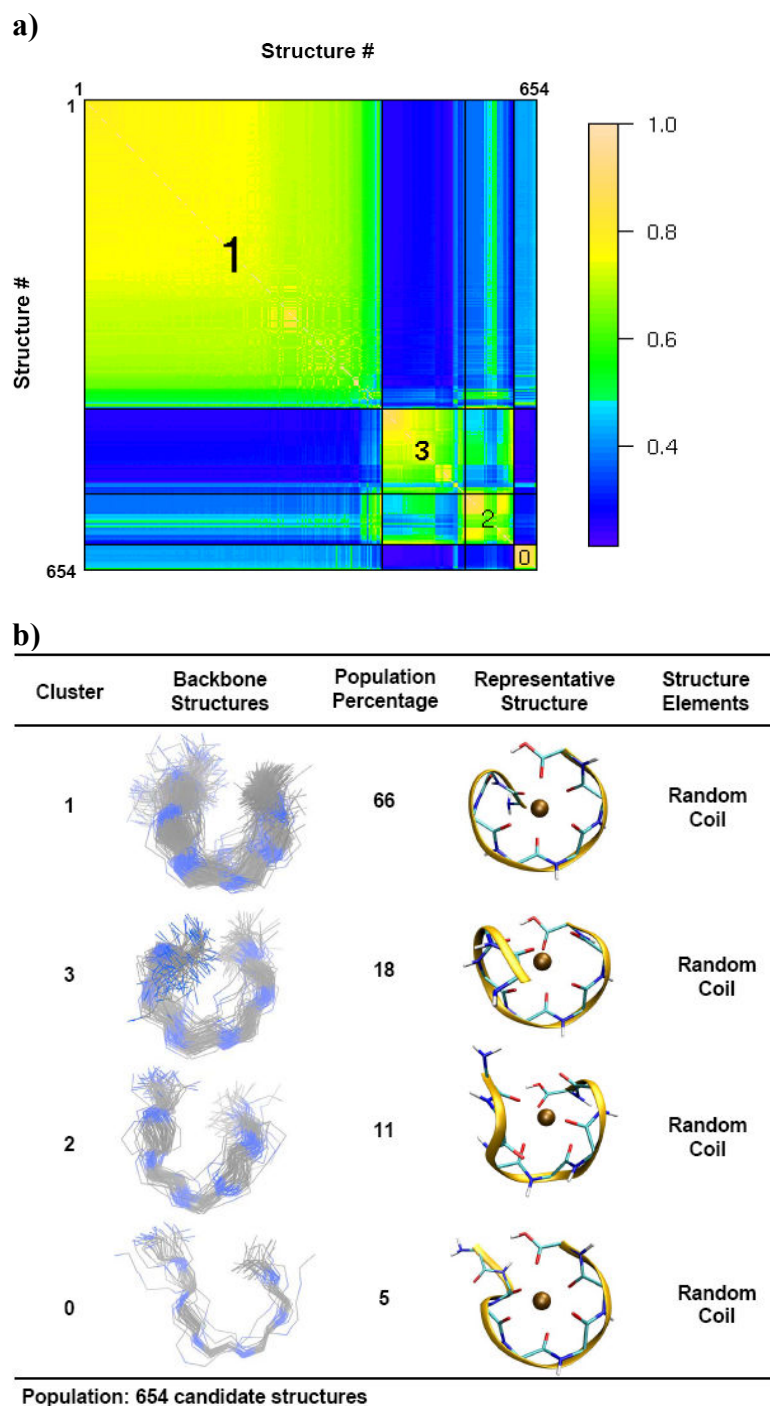
Figure 11. a) The pairwise probability plot of candidate structures for [MIFAGIK + Na]$^+$ ions formed by MALDI-IMMS, and b) the backbone structures of four most populated clusters for [MIFAGIK + Na]$^+$ ions.

a)

b)

Population: 1254 candidate structures
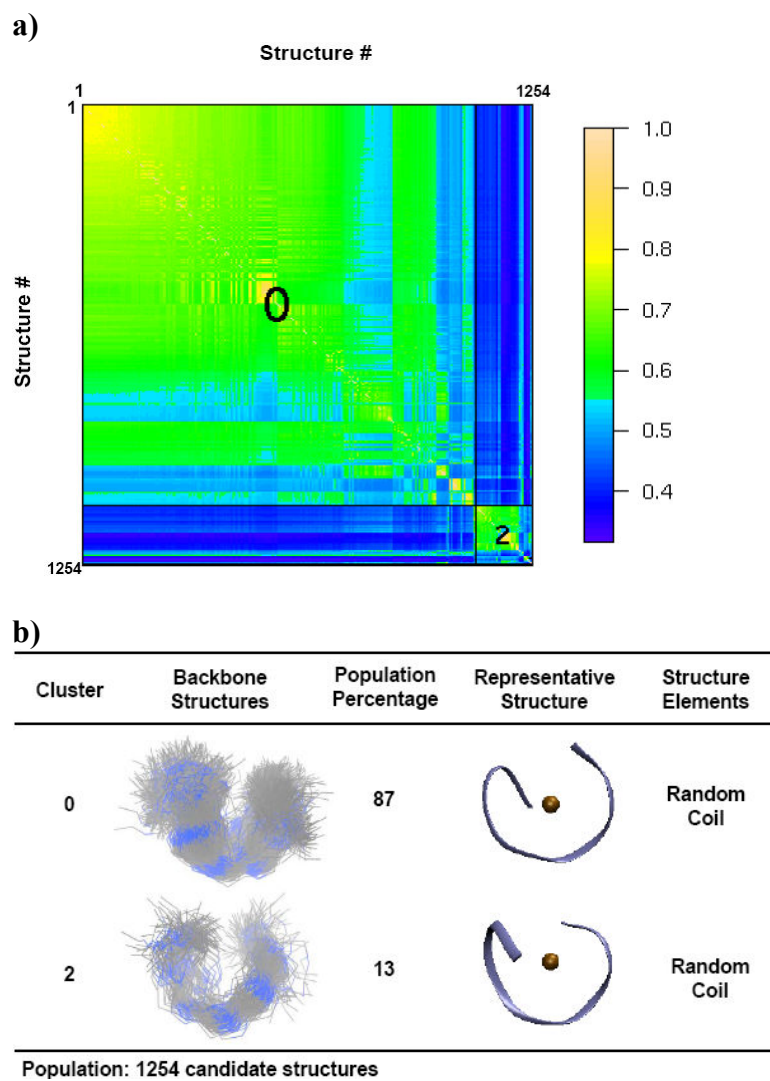
Figure 12. a) The pairwise probability plot of candidate structures for [*N-Ac*-MIFAGIK-*Ac* + Na]$^+$ ions formed by MALDI-IMMS, and b) the backbone structures of four most populated clusters for [*N-Ac*-MIFAGIK-*Ac* + Na]$^+$ ions.

a)

b)

Population: 1114 candidate structures
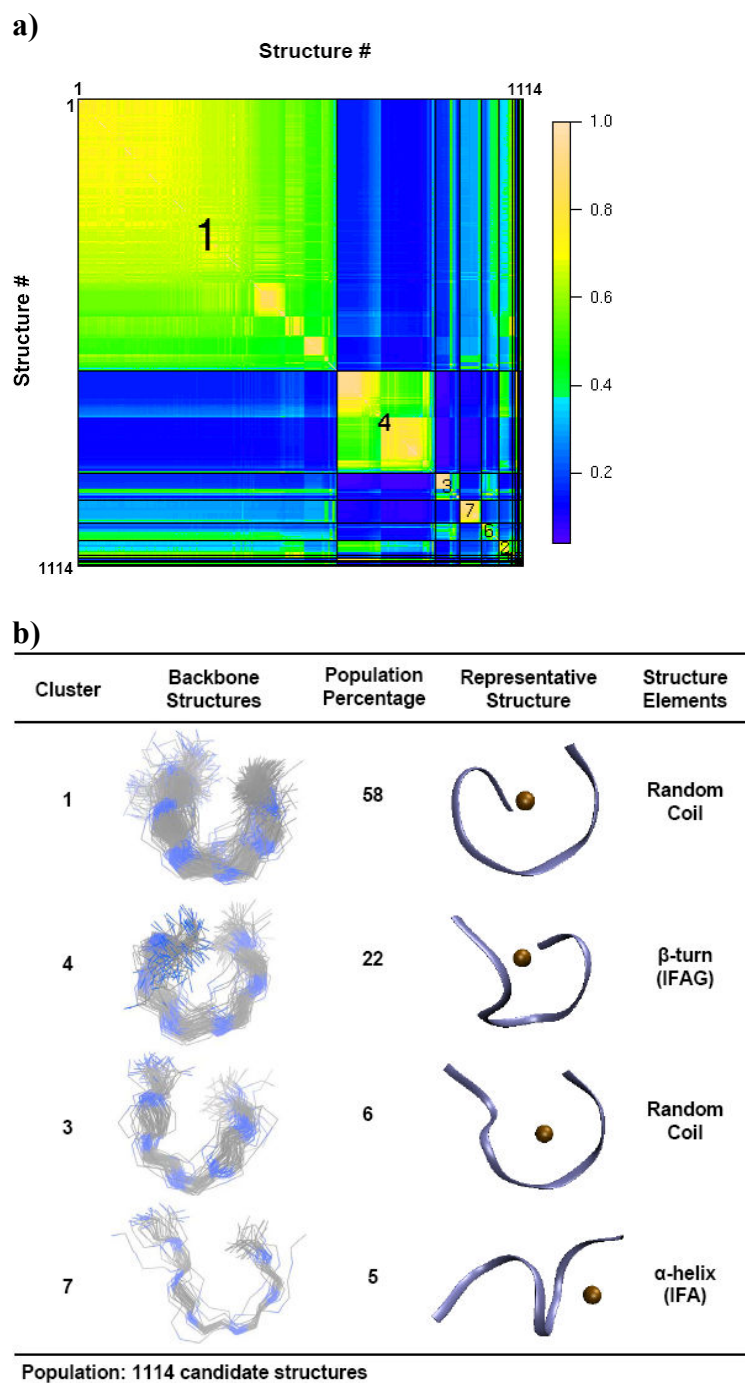
Figure 13. a) The pairwise probability plot of candidate structures for [MIFAGIK-*O-Me* + Na]$^+$ ions formed by MALDI-IMMS, and b) the backbone structures of four most populated clusters for [MIFAGIK-*O-Me* + Na]$^+$ ions.

N- and C-terminal derivatives of MIFAGIK

To test the 'sensitivity' of the cluster analysis method, we introduce subtle variations in peptide ion composition. Protecting groups which alter the polarity of the N- or C-terminus may influence the conformational preference of small peptides [90], and such changes are detected by using MD/cluster analysis for MIFAGIK derivatives [MIFAGIK-*OCH$_3$* + H]$^+$ and [*N-Ac*-MIFAGIK + H]$^+$ ions. The introduction of the methyl and acetyl groups should have minimal effects on ion structure because both groups are relatively small, steric effects should be small relative to charge-solvation by the amide backbone, Lys^7 is the preferred site of protonation, and the intramolecular interactions between the charged Lys^7 and the backbone carboxyl groups resembles those of MIFAGIK [M + H]$^+$ ions. The Ω_{meas} for [*N-Ac*-MIFAGIK + H]$^+$ and [MIFAGIK-*OCH$_3$* + H]$^+$ ions are 216 ± 4 Å^2 and 223 ± 4 Å^2, respectively. A total of 768 and 1,145 candidate structures that fall within $\pm2\%$ of the measured collision cross sections were subjected to cluster analysis. The results are shown in Figures 14 and 15. The [MIFAGIK-*OCH$_3$* + H]$^+$ ions adopt similar backbone conformations to MIFAGIK [M + H]$^+$ ions. This result is probably not surprising because the charge carrying site for the methyl ester derivative is the same as that of the underivatized peptide. Note, however, that differences in population size are detected. On the other hand, there are significant differences observed for the [*N-Ac*-MIFAGIK + H]$^+$ ions. For example, the most populated cluster corresponds to random coil structures rather than α-turn structures.

a)

b)

Cluster	Backbone Structures	Cluster Size	Representative Structure	Structure Elements
3		54		Random Coil
2		23		β-turn (IFAG)
0		12		α-helix (IFA)
4		9		Random Coil

Population: 1145 candidate structures
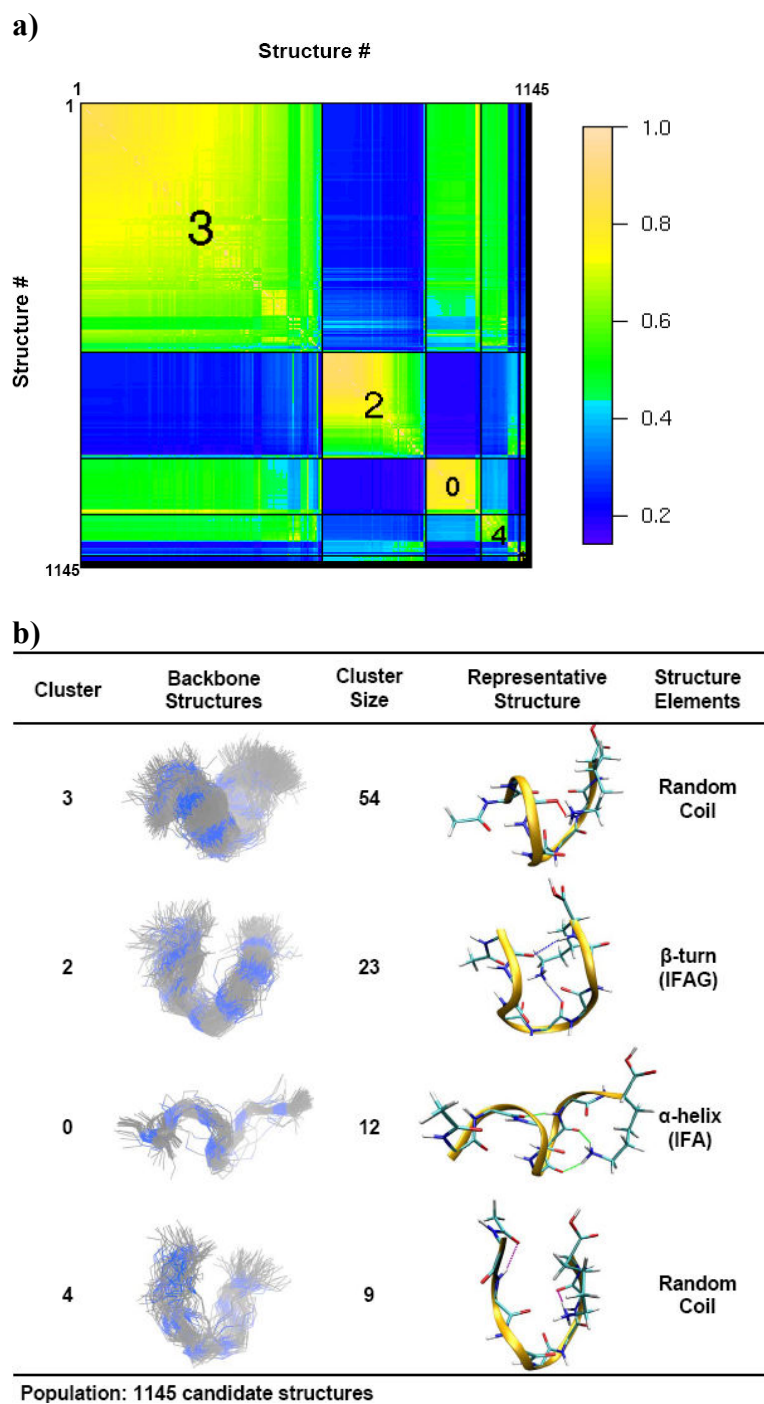
Figure 14. a) The pairwise probability plot of candidate structures for [*N-Ac*-MIFAGIK + H]$^{+}$ ions formed by MALDI-IMMS, and b) the backbone structures of four most populated clusters for [*N-Ac*-MIFAGIK + H]$^{+}$ ions.

a)

b)

Population: 768 candidate structures
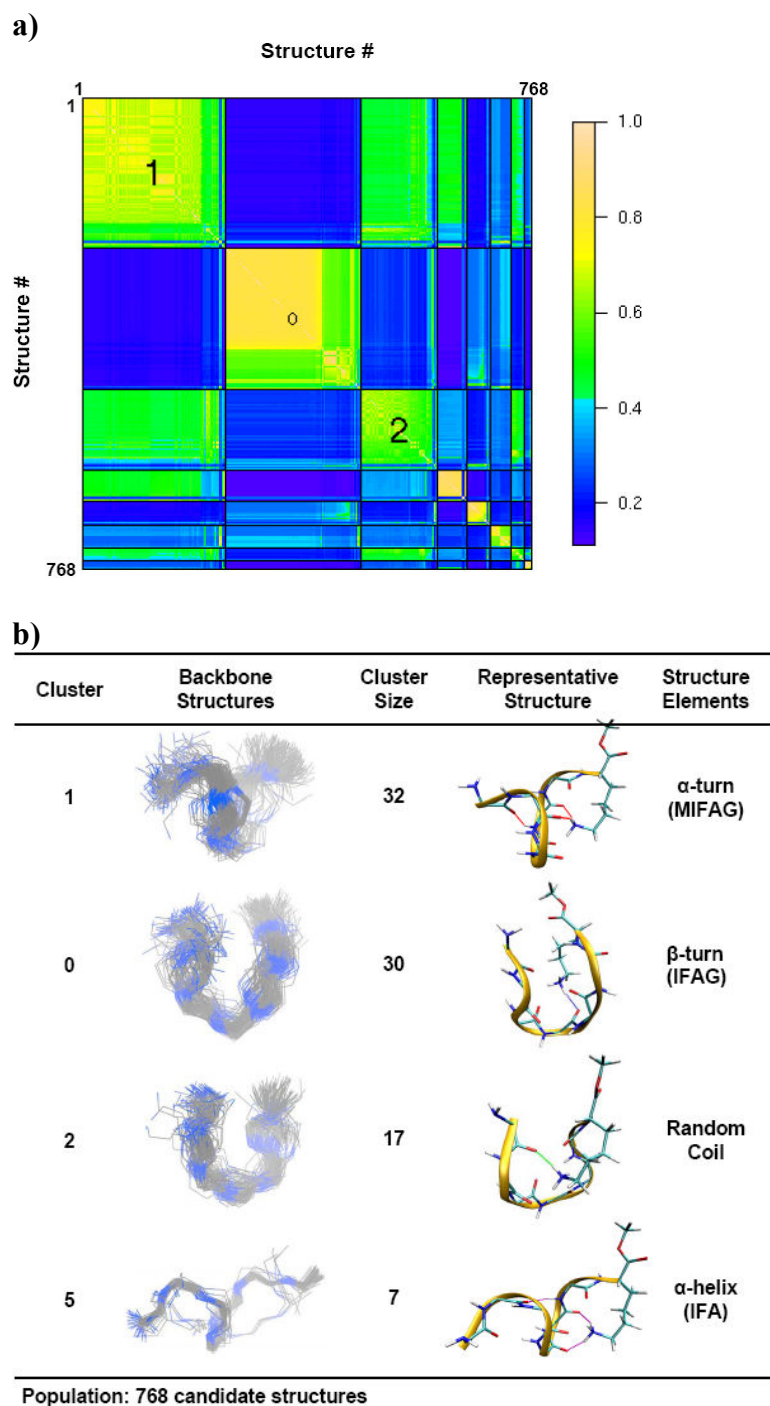
Figure 15. a) The pairwise probability plot of candidate structures for [MIFAGIK-*O-Me* + H]$^+$ ions formed by MALDI-IMMS, and b) the backbone structures of four most populated clusters for [MIFAGIK-*O-Me* + H]$^+$ ions.

MIFAGIK with bulky protecting group

To further validate the effect of charge site on the conformational preference of peptide MIFAGIK, [MIFAGIK-*N-Fmoc* + H]$^+$ and [*N-Fmoc*-MIFAGIK + H]$^+$ (Fmoc: 9-Fluorenylmethyloxycarbonyl) were produced by modified peptide MIFAGIK having the Fmoc protecting group on either N-terminus or Lys side residue. In the case of [*N-Fmoc*-MIFAGIK + H]$^+$, the charge site is on the side chain of Lys^7, whereas the proton is moved to the N-terminus for [MIFAGIK- *N-Fmoc* + H]$^+$ ions. Peptide sequences of two peptide ions are confirmed using tandem mass spectrometry and spectra are shown in Figure 16. As can be seen from the spectra, all the mass shifts are on y ions for [MIFAGIK-*N-Fmoc* + H]$^+$ ions and all the mass shifts are on b ions in the case of [*N-Fmoc*-MIFAGIK + H]$^+$.

The ATD plots for the [MIFAGIK-*N-Fmoc* + H]$^+$ and [*N-Fmoc*-MIFAGIK + H]$^+$ ions are shown in Figure 17. The centroid of ATD of [*N-Fmoc*-MIFAGIK + H]$^+$ is consistently *ca.* 8 μs longer than that for [MIFAGIK-*N-Fmoc* + H]$^+$ ions, and the calculated collision cross-sections for [*N-Fmoc*-MIFAGIK + H]$^+$ ions is *ca.* 3 Å^2 larger than that of [MIFAGIK-*N-Fmoc* + H]$^+$ ions. It is unclear whether these differences arise as a result of different structures or differences in the spatial projections of the Fmoc group, but more detailed studies on similar effects are currently underway.

Cluster analysis was performed on approximately 1000 candidate structures extracted from MD simulations for both peptide ions and results are illustrated in Figures 18 and 19.

a) [*N-Fmoc*-MIFAGIK + H]$^+$
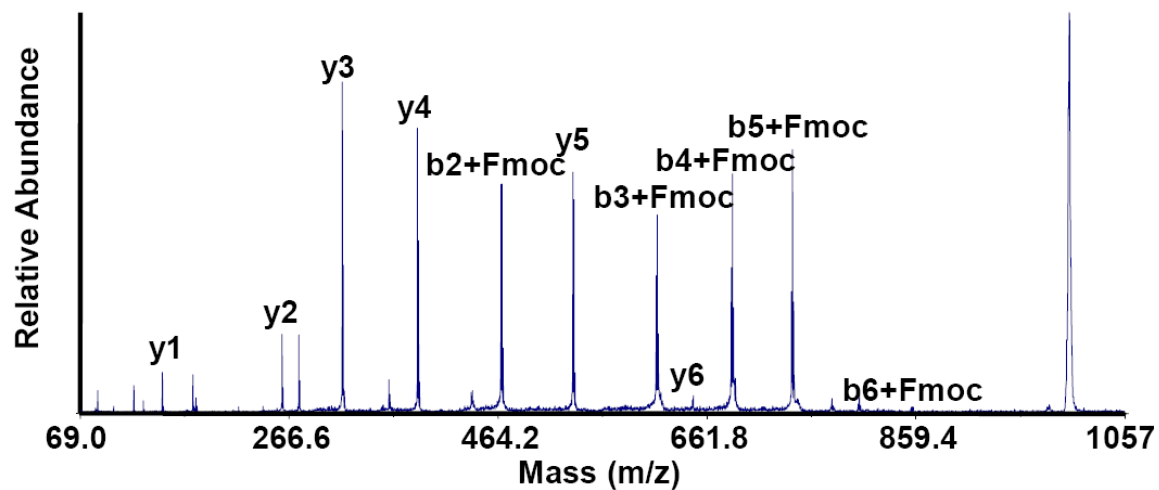
b) [MIFAGIK-*N-Fmoc* + H]$^+$
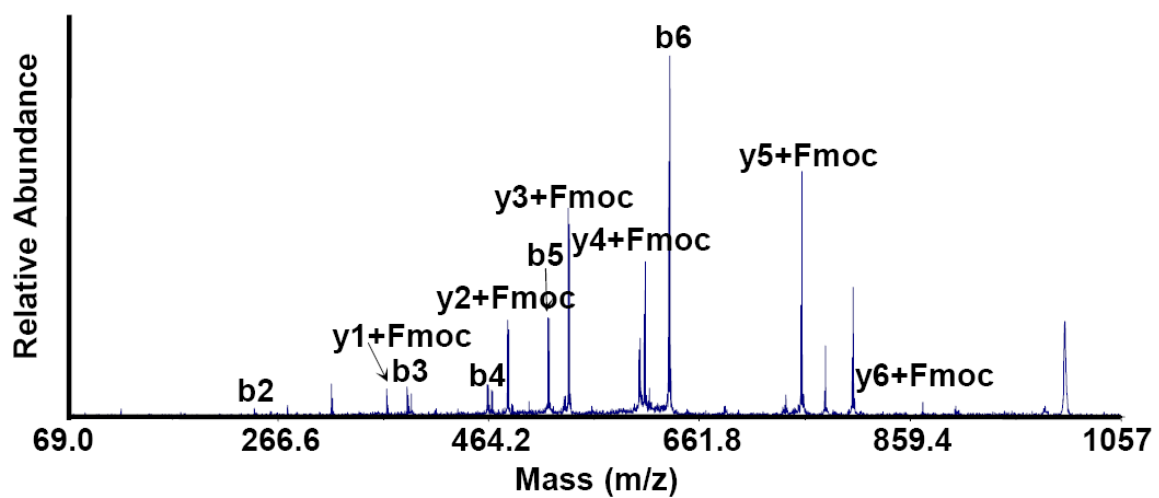
Figure 16. Tandem mass spectra of [*N-Fmoc*-MIFAGIK + H]$^+$ peptide ions and [MIFAGIK-*N-Fmoc* + H]$^+$ peptide ions.
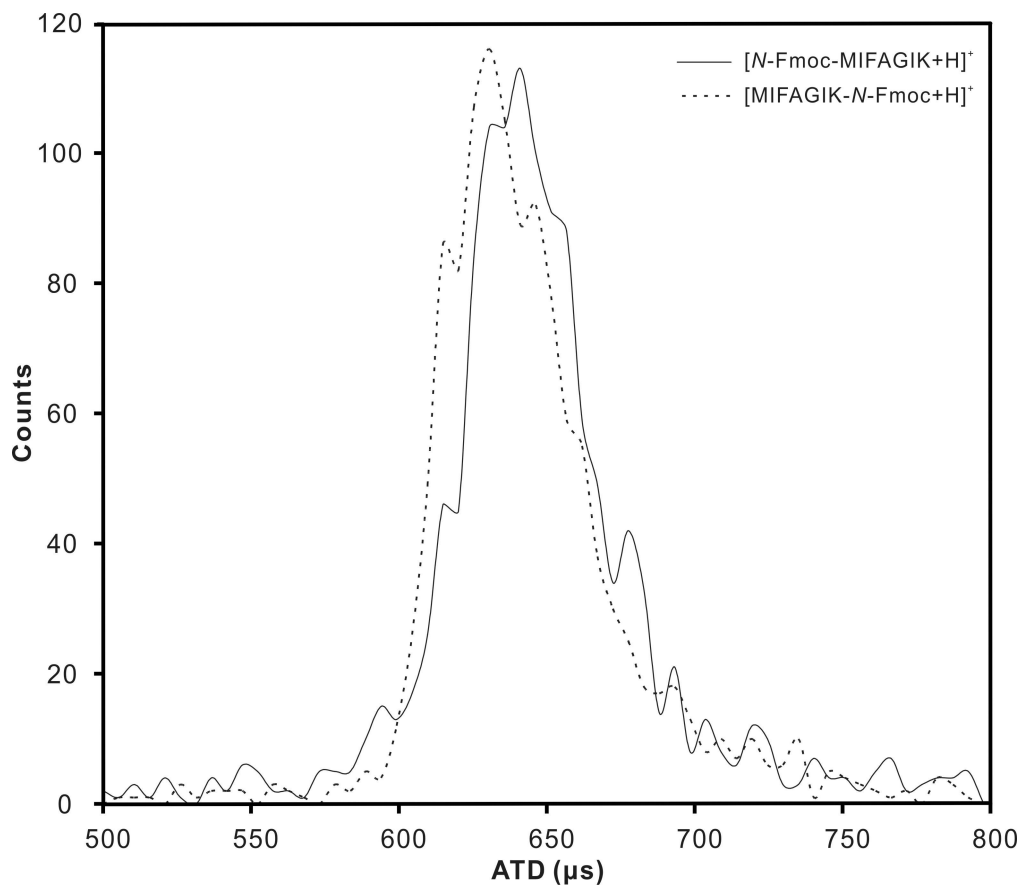
Figure 17. The ATD plot of [*N-Fmoc*-MIFAGIK + H]$^+$ (solid line) and [MIFAGIK-*N-Fmoc* + H]$^+$ (dotted line) peptide ions, the centroid ATD of [*N-Fmoc*-MIFAGIK + H]$^+$ is 642.4μs while the centroid ATD of [MIFAGIK-*N-Fmoc* + H]$^+$ is 635.3μs.

a)

Structure #

b)

Population: 1053 candidate structures
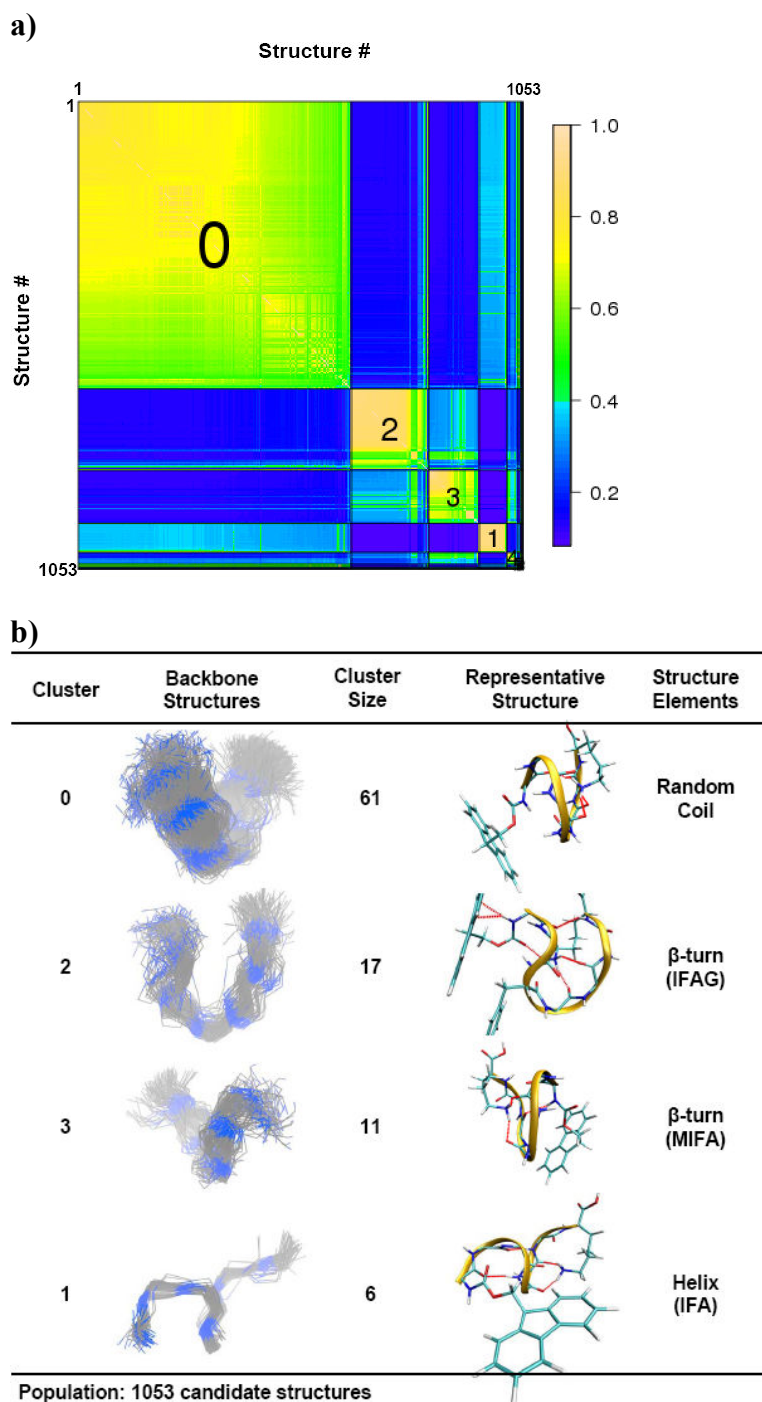
Figure 18. a) The pairwise probability plot of candidate structures for [*N-Fmoc*-MIFAGIK + H]⁺ ions formed by MALDI-IMMS, and b) the backbone structures of four most populated clusters for [*N-Fmoc*-MIFAGIK + H]⁺ ions.

a)

b)

Population: 1229 candidate structures
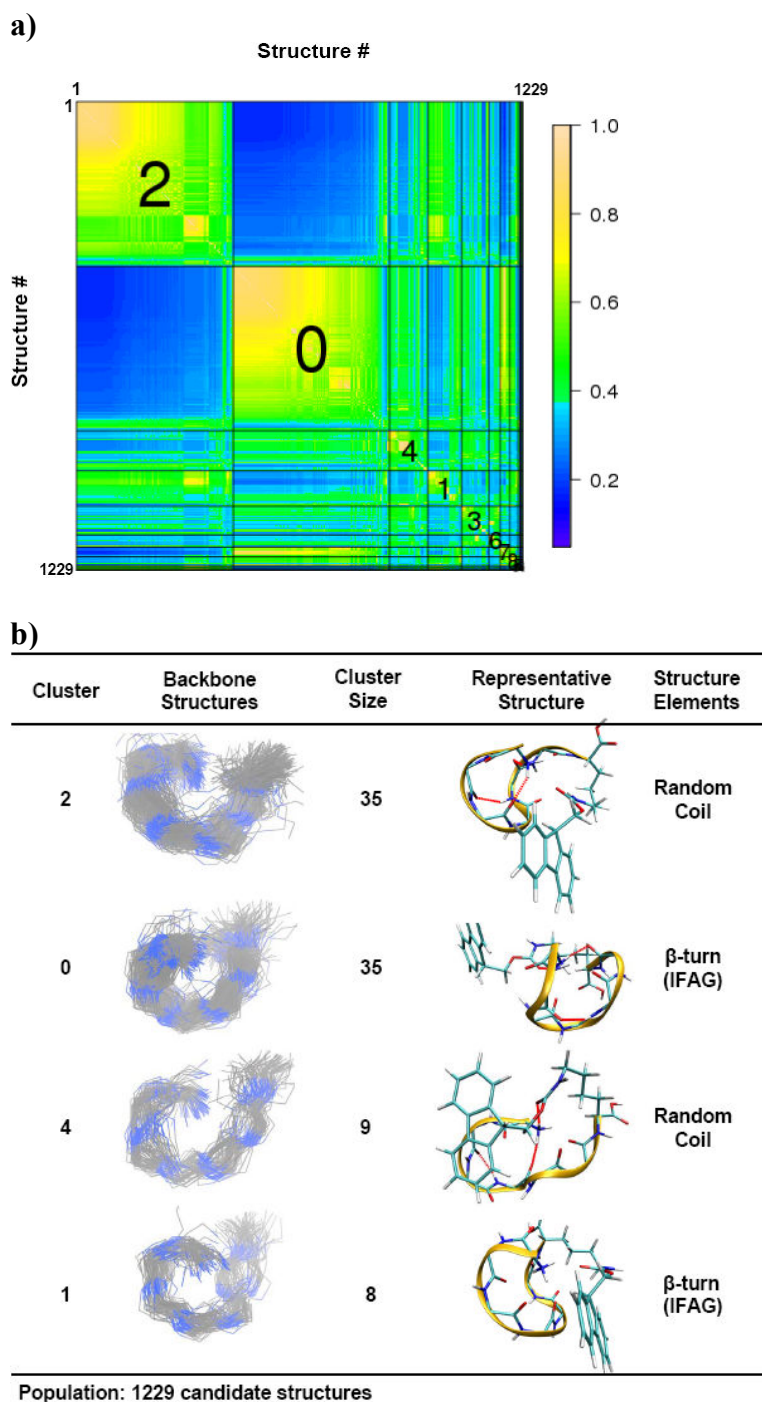
Figure 19. a) The pairwise probability plot of candidate structures for [MIFAGIK-*N-Fmoc* + H]$^+$ ions formed by MALDI-IMMS, and b) the backbone structures of four most populated clusters for [MIFAGIK-*N-Fmoc* + H]$^+$ ions.

For [*N-Fmoc*-MIFAGIK + H]$^+$, three clusters contain over 100 structures. The dominant conformer represents ~60% of the total ensemble, and this conformer appears to adopt a random coil structure, which is attributed to hydrogen bonding between the protonated Lys^7 and backbone amide groups. The remaining most populated clusters exhibits a turn in the Ile^2-Phe^3-Ala^4-Gly^5 sequence or a helical turn in Ile^2-Phe^3-Ala^4 sequence, which are similar to the most populated structures obtained for [MIFAGIK + H]$^+$ ions. In the case of [MIFAGIK-*N-Fmoc* + H]$^+$ ions, the two most populated clusters account for over 70% of the total ensemble. Interestingly, both conformations contain a bridge between Ile^2 and Gly^5 which is quite different from that for peptides with the charge on Lys^7. It is interesting to note that helical conformations are not found for this ion since the preferred charge site of [MIFAGIK-*N-Fmoc* + H] is probably the N-terminus and the helix is probably destablized by the interaction between the charge and the helix macrodipole.

Conclusion

Cluster analysis of IM-MS data is a powerful approach for classification of peptide structures derived from molecular dynamics simulations. Each cluster represents probable conformations observed for gas-phase ions. Specifically for the peptide MIFAGIK, cluster analysis suggests that [M + H]$^+$ ions formed by MALDI or ESI both prefer a α-turn structure; this structural preference is probably the result of stabilization afforded by charge-solvation by the backbone amide groups. That is, there appears to be a significance preference for helical or partial helical conformers by the low energy ions,

but the major fraction of the ion populations exists as α-turn, random coil, and β-turn conformers.   Although the derivatives of MIFAGIK which influence charge site and intramolecular interactions alter the distribution of various conformers, the most dramatic changes are observed for the $[M + Na]^+$ ion, which show a strong preference for random coil conformers owing to the strong solvation by the backbone amide groups. We tested the 'sensitivity' of the cluster analysis method by introducing subtle variations in peptide ion composition as well.   For example, the N-acetyl and methylester derivatives as well as the Boc and Fmoc derivations all share very similar backbone structures.   All the results presented here support the hypothesis that difference in collision cross-sections for MIFAGIK peptide ions are related to the different orientation of side chains, *i.e.,* MIFAGIK is a small peptide which has limited degrees-of-freedom for positioning backbone structural elements.   We are currently expanding these studies to studies of the results to larger peptide systems in an effort to increase statistical confidence of gas-phase conformation assignment.

CHAPTER IV

ION MOBILITY-MASS SPECTROMETRY OF

OLIGODEOXYNUCLEOTIDES

Introduction

Structure of DNA

Deoxyribonucleic acids (DNA) are polymers of nucleotides (nt) that contain the genetic information used in the functioning of all living organisms. Nucleotides are composed of a heterocyclic base, a sugar and a phosphate group. In DNA, the sugar unit is 2'-deoxyribose, the five-carbon ring, joined together through phosphor-ester bonds between the third (C3') and fifth carbon atoms (C5') of adjacent sugar rings. The heterocyclic bases found in DNA are structural derivatives of either purine (adenine (A) and guanine (G)) or pyrimidine (cytosine (C) and thymine (T)). They attach to the C1' of deoxyribose through N9 of the purine bases or N1 of the pyrimidine bases with an N-glycosidic bond.

DNA is a double helical structure first determined by James Watson and Francis Crick in 1953 [91]. The most biologically common form of DNA is a right-handed double helix known as B-DNA. The structure consists of two antiparallel polynucleotide strands with the base is hydrogen bonded to the base in the opposite strand to a planar base pair. There are only two types of base pairs (Watson-Crick base pair) with A bonding only to T, and C bonding only to G. The "ideal" B-DNA helix has 10 base pairs per turn and a pitch (rise per turn) of 34 Å. Two major structural variants of DNA are A-DNA and Z-DNA. Compared to B-DNA, A-DNA form is a wider and flatter

right-handed helix, with a shallow, wide minor groove and a narrower, deeper major groove. A-DNA form occurs under non-physiological conditions in partially dehydrated samples of DNA. Specific segments of B-DNA where the bases have been chemically modified by methylation may undergo a larger change in conformation and convert to Z-DNA form. The Z conformation is a left-handed double helix which has twelve Watson-Crick base pairs per turn, a pitch of 45 Å, a deep minor groove, and no discernible major groove. The unusual structure features can be recognized by specific Z-DNA binding proteins and may be involved in the regulation of transcription [92-94].

The conformational flexibility of DNA is greatly restricted by a variety of internal constraints. Six torsion angles of the sugar-phosphate backbone and one torsion angle between the base and the glycosidic bond define the conformation of each nucleotide unit. The rotation of a base around its glycosidic bond is sterically hindered. It is found that all bases are in the anti conformation in most double-helical DNA. The flexibility of deoxyribose ring is also limited their cyclic structures. The two most common ribose conformations are known as C2'-endo and C3'-endo. The sugar ring pucker is important in DNA conformation because it governs the relative orientations of the phosphate groups to each ribose residue. The C2'-endo conformation occurs in B-DNA, whereas A-DNA is C3'-endo. In Z-DNA, the purines are all C3'-endo and the pyrimidine are C2'-endo. The sugar-phosphate backbone chains of DNA are stiff as well. The torsion angles are quite restricted because of noncovalent interactions between the ribose ring and the phosphate groups and steric interference between residues. These properties

make the N-glycosyl bond a key element in understanding the possible conformations of nucleotides and and, subsequently, of nucleic acids.

Complementary base pairing is the most important features of DNA to form double-stranded DNA conformation, Base stacking (interaction between the π-electron systems of the bases), hydrophobic interactions between adjacent base pairs and electrostatic interactions of the charged phosphate groups also contribute a lot to the stabilities of DNA structures in aqueous solution. The strength of the double-stranded interacting structure increases with an increasing number of the base pairs, and the strand with higher number of GC base pairs is more stable owing to the contribution of π-stacking interactions. The strength of such interaction can be measured by the melting temperature (T_m) [95].

Single-stranded DNA is rarely occurs under physiological conditions. The base pairing between complementary sequences within the single strand makes the polynucleotide to fold back on itself, giving rise to stem-and-loop structures or more complex structures.

Acid-base chemistry

Nucleic acids have several functional groups that act as acids and bases, *i.e.*, phosphate group and carbonyl groups, etc. The structures and reactivities of DNA molecules are greatly affected by the ambient pH. An appreciation of acid-base chemistry is therefore essential for understanding the biological roles of DNA molecules. Nucleobases are weak bases, the order of basicity is C > A > G as demonstrated by their

pKa values of the protonated species. T exhibits no basic function. Actually, the basicity of a nucleobase residue can vary easily due to the neighboring function groups. The pKa values for nucleosides and nucleotides are ionic strength [96] and temperature dependent[97], so they differ considerably in the literature. The effect of ionic is pronounced in the nucleotides due to the influence of salts on the phosphate group ionization, which in turn affects the acid-base properties of the base moiety. Negatively charged phosphate groups favor protonation of the base and formation of zwitterion, raising the basicity of A, C, and G, suppressing the formation of negatively charged base residues, and reducing the acidity of T. For the influence of the temperature on the ionization of base, suppression can occur at basic condition for experiments using increased temperatures.

The most likely protonated sites within the nucleobases are N3 for C, N1 for A, and N7 for G [98]. Normally, oligonucleotides are present as salts countered by cations such as Na^+, K^+, or ammonium ions. In basic and neutral solution, oligonucleotides are present as oligoanions with deprotonation of the nucleobases. Under acidic condition, mixtures of zwitterions (deprotonated phosphate groups and protonated nucleobases) are observed; the charge location is dependent on the base composition of the oligonucleotides. Protonation of the phosphate groups can occur at even lower pH solution, resulting positively charged nucleotides.

The proton affinities (PA) of gas phase oligonucleotides have been examined for studying ion formation of oligonucleotides in mass spectrometry [99, 100]. Greco et al. proposed the PA of the different neutral deoxyribonucleoside monophosphate (dNMP)

increase in the order T (224.9 kcal/mol) << C (233. kcal/mol) < A (233.6 kcal/mol) < G (234.4 kcal/mol) [101]. And the PA of the phosphate group is predicted ~ 220 kcal/mol by Rodgers et al [100]. These values define the charge location of oligonucleotide ions. In the case of protonated olignucleotides, the charge-giving proton will be expected at the nucleobase, while for negatively charged oligonucleotide ions, deprotonation of phosphate groups may be the preferred structure of the analyte ions.

Mass spectrometry (MS) of DNA

Mass spectrometry has played an important role in the analysis of nucleic acids for years. With the advent of ESI and MALDI, the oligonucleotides are able to be analyzed at tens of femtomoles. ESI and MALDI typically generate prtonated and/or deprotonated oligonucleotide ions of $[M + nH]^{n+}$ and $[M - nH]^{n-}$, respectively. Cationized species $[M + nNa/K]^{n+}$ and multimers, e.g., $[2M + nH]^{n+}$, are observed as well. With ESI, higher charge states are formed preferably, depending on the molecular mass of the analyte, the solvent conditions, and ion-source conditions. In MALDI-MS, generally singly charged molecular ions dominate the spectra. Sometimes doubly or triply charged species are also observed with particular matrix. In ESI-MS of oligonucleotides, negatively charged ions yield better results than positively charged ions. For MALDI-MS, both ion detection modes can work well with the right choice of matrix. However, there are several difficulties in the mass analysis of oligonucleotides. One is the formation of alkali metal adducts in the type of $[M - (n+m)H +mNa/K]^{n-}$. Such distribution of the molecular ions not only leads to a reduction of the signal-to-

noise (S/N) ratio, but also requires a high instrumental resolution power to separate the individual ions. If they are not separated and merge into one broad peak, it can cause the loss of mass accuracy. Furthermore, mixtures containing with small mass differences will not be separated. This problem can be solved by adding ammonium ions as counter ions to generate the free acids form of the oligonucleotide ions [102]. Another issue in the analysis of oligonucleotides is the extensive fragmentation of the molecular ions. The oligonucleotides tend to lose their nucleic bases by a 1,2-elimination mechanism, followed by multiple cleavages of the sugar-phosphate backbone [103-105]. The problem is more pronounced in MALDI-MS than in ESI-MS. The generation of intact molecular ions of oligonucleotides above ten base units in length causes more difficulties with MALDI than with ESI.

In MALDI-MS, the success of the analysis of oligonucleotides is strongly dependent on the choice of the matrix and the preparation of the sample. In the first MALDI studies of oligonucleotides, only matices that had proven to be useful for the analysis of peptides were examined such as nicotinic acid, ferulic acid, sinapic acid, and 2,5-dihydroxybenzoic acid (2,5-DHB) at a laser wavelength of 337 (nitrogen laser) or 355 nm (frequency-tripled Nd:YAG laser) [106-110]. It was found that these matrices were not good choice for the analysis of oligonucleotides, and, therefore, new matrices were introduced specific for the oligonucleotides by several research groups and major improvement was achieved for the analysis of oligonucleotides, such as 3-hydroxypicolinic acid (3-HPA) [111], 3-aminopicolinic acid [112], 2',4',6'-trihydroxy acetophenone (2',4',6'-THAP) [113], and 6-aza-2-thiothymine (ATT) [114]. Among all

those matrices, the discovery of 3-HPA gives the major improvement. With 3-HPA as the matrix, a significant reduction of ion fragmentation can be achieved, which allows the analysis of larger oligonucleotides and yields a better S/N ratio. Furthermore, 3-HPA is suited for most frequently used laser on MALDI-MS, such as nitrogen laser or YAG laser. It thus became the standard matrix for the analysis of oligonucleotides. Combined with picolinic acid (PA) and ammonium citrate, 3-HPA is the preferred matrix to measure larger nucleic acids.

Several strategies for sample preparation have been used for the suppression of metal-cation adduction. A simple and efficient say to replace metal ions is the addition of NH_4^+-loaded cation exchange beads to the final sample droplet. After the solvent evaporates, the beads accumulate in the center of the sample from where they can be easily removed before the mass analysis [115, 116]. Alternative approach of the cation exchange beads is the use of Naflon [117] or NC-coated support [118] for ontarget cation exchange, they proven to be efficient in suppression of metal cation adduction. Another approach is the addition of ammonium salt to the sample to eliminate the interference from metal ions [113].

The response of oligonucleotides to MALDI-MS depends on the nature, size and sequence composition of nucleic acids as well. Oligonucleotides containing only T give a much stronger signal than those composed of other bases. T is much harder to be protonated than other bases and get lost during the fragmentation, which results in better stability of the polymer chain. When the size of the nucleotide increases it also becomes harder to be ionized and detected. In addition, the chance of forming metal adducts and

fragmentation also increases with increasing analyte size. These all attribute to the low S/N ratio and the low resolution of the spectra for the larger oligonucleotides [119]. The error on the measured molecular mass obtained by MALDI varies between 0.01 to 0.05% for small oligonucleotides. The error increases above 0.5% for large nucleotides. In deed, at higher mass range, the signal of metal adducts or the fragmentation ions of oligonucleotides are more abundant. The peaks are broad and cannot be resolved.

Most of MALDI-MS has been used for the analysis of covalent complexes. However, a few observations of noncovalent interactions have been reported. The first mass spectra of intact double-stranded DNA was reported by Lecchi et al. using MALDI-TOF-MS with ATT/ammonium citrate as the matrix at a laser wavelength of 337 nm [120]. The EcoRI adaptor 12/16, containing 12 Waston-Crick base pairs, was used as sample; both single strands and the complementary double strand were detected, but no homodimer of either of the single strands was observed. The authors have demonstrated that is not nonspecific dimer formation by digesting with nuclease S1, an enzyme that preferentially cleaves single stranded DNA. No DNA duplex was not detected when using other matrices commonly employed for oligonucleotide analysis.

Ion mobility mass spectrometry (IM-MS) of DNA

Ion mobility mass spectrometry (IM-MS) has been used to explore a variety of bio-molecular systems ranging from high-throughput separation [34, 65] to detailed examination of biomolecule conformations in the gas phase [56, 121, 122]. Ion mobility separations are performed based on the ion-neutral collision cross-section with buffer

gas under a weak electric field, which is proportional to apparent ion surface area. When coupled with mass spectrometry, it provides 2D separations both in the dimension of the structural size and the dimension of mass-to-charge (m/z). IM-MS separation is highly selective, it can distinguish structural isomers and different classes of molecules (*i.e.*, peptide ions from DNA or carbon cluster ion or chemical noise) based on their different signal location in 2D IM-MS plots [43, 44] The collision cross-section data can be transformed to structural information of ions using molecular dynamics methods. Of the various methods used to study gas phase biomolecules, measurements of collision cross sections give the most direct insights to folding.

A lot of research has been done on developing IM-MS for biophysical studies of peptides and proteins [34, 52, 55, 56, 63, 64]. However, there is no much work has been reported on the IM-MS applications to DNA. The problem is that the oligonucleotides tend to fragment, especially for the larger ones, which limits the size of oligonucleotide that can be studied by IM-MS. Koomen et al.[43] reported several applications of MALDI-IM-TOFMS on oligonucleotides, including separation of protonated oligonucleotide mixtures generated by MALDI, differentiation between DNA and peptide ion signals, oligonucleotide sequencing, and analysis of chemically modified DNA. With a mobility resolution of 20 ~ 30, MALDI-IM-TOFMS was used to separate oligonucleotides of different length, but not to differentiate between isomers or even different compositions of the same length. It was also used to separate the mixtures of oligonucleotides and peptides and resolve DNA-pltinum adducts from the corresponding unmodified oligonucleotides. Oligonucleotide sequencing was also possible by

MALDI-IM-MS. Gidden et al. reported the studies on gas phase conformations of deprotonated dinucleotide (dGT⁻, and dTG⁻) [123] and trinucleotides (dGTT⁻, dTGT⁻, and dTTG⁻) [124] using MALDI-IM-MS combined with molecular dynamics calculations. The results indicate that both "open" and "folded" conformations were determined for each case. The different conformers can only be resolved at 80 K, while at 300 K; they rapidly interconvert and yield a single, time-average conformation with a cross section between the values predicted for the folded and open conformers. The isomeriztion barrier between different conformers for each case was calculated as well at a range from 0 ~ 5 kcal/mol.

The goal of this study is to use ion mobility mass spectrometry methods to examine the gas phase conformations of oligonucleotides. A series of oligonucleotide of different length will be tested by MALDI-IM-TOFMS to define the conformational space of oligonucleotide. With a novel non-Boltzman sampling molecular dynamics (MD) on model oliogonucleotide ions, a more accurate description of the gas phase conformational space of oligonucleotide ions can be achieved, which involving the thermodynamically accessible structures that correspond to real experimental condition.

Experimental

Samples and preparation

The DNA oligonucleotides were purchased from Invitrogen Corp. (Carlsbad, CA) (Table 8). They are: dGGATC (mass 1502), dCTAATC (mass 1750), dCTGGTC (mass 1782), dGATTAG (mass 1830), dGACCAG (mass 1800), dGATTAGCA (mass 2432), d

Table 8. A list of all the oligonucleotides sequences examined in this study and the molecular weights of their monomer form.

nt	Sequence	MW	[M+H]$^+$	[M-H]$^-$
5	GGATC	1502.30	1503.31	1501.29
6	CTAATC	1750.34	1751.35	1749.33
6	GATTAG	1830.35	1831.36	1829.35
6	CTGGTC	1782.33	1783.34	1781.32
6	GACCAG	1800.36	1801.36	1799.35
8	CTCTAGAG	2408.45	2409.45	2407.44
8	GATTAGCA	2432.46	2433.47	2431.45
9	GACCCTGGA	2722.50	2723.51	2721.49
10	CACATGGCTG	3026.54	3027.55	3025.54
10	ATCGATCGAT	3025.55	3026.56	3024.54
11	CAGCCATGTGA	3339.60	3340.61	3338.59
12	GAGACCTTAGCA	3652.66	3653.67	3651.65
13	GCACATGGCTGAC	3957.70	3958.71	3956.69
14	CTGCAGCCATGTGA	4261.75	4262.75	4260.74

CTCTAGAG (mass 2408), dGACCCTGGA (mass 2722), dCACATGGCTG (mass 3026), dATCGATCGAT (mass 3025), dCAGCCATGTGA (mass 3340), dGAGACCTTAGCA (mass 3653), dGCACATGGCTGAC (mass 3958), and dCTGCAGCCATGTGA (mass 4262). All matrixes and matrix additives are purchased from Sigma (St. Louis, MO) and Aldrich (Milwaukee, WI). The matrixes include 2, 4, 6-trihydroxyacetophenone (THAP), α-cyano-4-hydroxycinnamic acid (αCHCA), 2,5-dihyroxybenzoic acid (DHB), and 3-hydroxypicolinic acid (3-HPA). The matrix additives are restricted to diammonium hydrogencitrate (DAC), and triammonium citrate (AC).

Samples are prepared by both dried droplet methods and overlayer methods. Table 9 lists all of the preparation methods that are examined. The matrix solutions are prepared fresh daily and mixed 1:1:1 with matrix additive (10mg/mL diammonium hydrogencitrate and 50 mg/mL triammonium citrate) and oligodeoxynucleotide stanards (1 mg/mL in distilled deionized water). The matrix-to-analyte ratios are in the range of 40:1−2,000:1.

Dried droplet preparations are made using the following matrix solutions: 35 mg/mL 3-hydroxypicolinic acid in 50% acetontrile, or 20 mg/mL 2,4,6-trihydroxyacetophenone in 100% HPLC grade methonal, or 50 mg/mL 2,4,6-trihydroxyacetophenone in 100% distilled deionized water, or 40 mg/mL 2,5-dihydroxybenzoic acid in 85% aqueous diammonium hydrogencitrate (10−20 mg/mL) and 15% acetonitrile.

Table 9. A list of seven sample preparation protocols examined in this study.

Method	No.	Matrix	Co-Matrix	Matrix additive
Dried droplet	1	HPA (35mg/mL 50%ACN)		DAC (10mg/mL100%ddH$_2$O)
	2	THAP (20mg/mL 100%MeOH)		
	3	THAP (50mg/mL 100%ddH$_2$O)		AC (50mg/mL 100%H$_2$O)
	4	DHB (40mg/mL 15:85 ACN:DAC 10mg/mL ddH$_2$O)		
	5	DHB (100mg/mL 100%MeOH)		
Overlayer	6	THAP (30mg/ml in 50%MeOH)	αCHCA (30mg/ml in 100%MeOH)	DAC (10mg/mL100%ddH$_2$O)
	7	αCHCA (30mg/mL 100%MeOH)	αCHCA (30mg/ml in 100%MeOH)	DAC (10mg/mL100%ddH$_2$O)

Overlayer methods are made using the following matrix solutions: 30 mg/mL 2, 4, 6-trihydroxyacetophenone in 50% HPLC grade methonal, or 30 mg/mL α-cyano-4-hydroxycinnamic acid in 100% HPLC grade methonal. The mixture of matrix and analyte solution 0.5 μL is then deposited on a bed of matrix deposited previously from concentrated co-matrix solution (30 mg/mL α-cyano-4-hydroxycinnamic acid in 100% HPLC grade methonal).

MALDI-TOFMS

All MALDI-TOFMS measurements are preformed by both a Perseptive Biosystems Voyager Elite XL (STR) (PE Biosystems, Framingham, MA) and Applied Biosystems 4700 Proteomics analyzer (Applied Biosystems, Forster City, CA). Reflected mode acquisition is used to monitor both negative and positive ions. STR is equipped with delayed extraction and a nitrogen laser. Accelerating voltage is set at either 20 kV in negative ion mode or 25 kV in positive ion mode, the grid voltage is 69−71% of the acceleration voltage, and the guide wire is set to 0.010−0.025% of the acceleration voltage. Resolution and mass accuracy are comparable for all preparation methods carried out in both instruments; signal-to-noise ratios are higher in positive ion mode, so it is used primarily.

MALDI-IM-TOFMS

Stock solutions of peptide (1mg/mL) were mixed with re-crystallized α-cyano-4-hydroxycinnamic acid (Sigma) in a molar ratio of ~2000:1, and 1 μl aliquots of the peptide/matrix mixture was deposited on a stainless steel MALDI sample plate.

The MALDI-IM-TOFMS analysis was performed using a home-built instrument previously described [68]. Briefly, MALDI was performed using a frequency tripled (355 nm) Nd:YAG laser (CrystaLaser, Reno, NV) operated at a pulse rate of 200-400 Hz. The resulting ions were introduced into a 15 cm drift tube maintained at a pressure of ca. 3.0 Torr of He maintained at room temperature. Under these conditions the E/p ratios range from 18 to 34 Vcm^{-1}torr^{-1}, which provides mobility resolution of 15 to 25. The ions eluting the drift cell were extracted and mass analyzed by an orthogonal reflectron-TOF, typical m/z resolution ca. 2000 - 4000. The mass spectrometer was externally calibrated using two-point calibration on C_{60} (M_r = 720) and C_{70} (M_r = 840) radical cations (Sigma) [5]. The measurements of collision cross sections were externally calibrated with $[M + H]^+$ ions of bradykinin (Ω_{meas} = 245 Å^2) and substance P (Ω_{meas} = 292 Å^2) [35]. The 2D IM-MS data was acquired and processed by using custom software (Ionwerks, Inc.).

The experimental ion-neutral collision cross-sections were determined from the empirical drift times (t_d) by the following equation [69]:

$$\Omega_{meas} = \frac{(18\pi)^{\frac{1}{2}}}{16} \frac{ze}{(k_b T)^{\frac{1}{2}}} \left(\frac{1}{m_I} + \frac{1}{m_B} \right)^{\frac{1}{2}} \frac{t_d E}{L} \frac{760}{P} \frac{T}{273.2} \frac{1}{N_0} \tag{1}$$

where z is the charge of the ion, e is elementary charge, N_0 is the number density of the drift gas at STP, k_b is Boltzmann's constant, m_I is the mass of the ion, m_B is the mass of buffer gas, E is the IM electric field strength, L is the drift tube length, P is the buffer gas pressure and T is the system temperature.

Charge derivation of DNA for AMBER

A new force field topology database of RESP atomic charge values was developed for modeling DNA. RESP charge derivation for new nucleotide fragments was performed using Dimethylpohosphate (conformation gauche, gauche), neutralized Dimethylphate (conformation gauche, gauche), the 4 neutral DNA nucleosides [Deoxyadenosine, Deoxycytidine, Deoxyguanosine and Deoxythymidine] (conformations C2'endo and C3'endo), and the 3 protonated DNA nucleosides [Deoxyadenosine, Deoxycytidine and Deoxyguanosine] (conformations C2'endo and C3'endo). All geometries were optimized using the HF/6-31G* theory level and four molecular orientations for each optimized geometry were involved in the charge fitting procedure to yield reproducible atom charge values. Two inter-molecular charge constraints between the methyl group of Dimethylpohosphate and the HO3' and HO5' hydroxyls of the target nucleotides were used during the fitting step allowing the definition of the required molecular fragments. Inter-molecular charge equivalencing between the nucleosides were used as well to make the charge values of the deoxyribose atoms (exluding the C1' and H1' atoms) equivalent. The charge derivation procedure was automatically carried out using the R.E.D-IV program [125] and the central, 5'-terminal and 3'-terminal fragments of a nucleotide were simultaneously generated in a single charge derivation. Both topologies A and B, which present the phosphate group resided either at the position 5' or 3' of the target nucleotides, were obtained in this project. The new charge values are compatible with the Cornell *et al.* AMBER force

field and proved to be highly reproducible. All data are available at http://q4md-forcefieldtools.org/REDDB/projects/F-83/.

Enhanced sampling molecular dynamics (ES-MD)

The method we used here is based on the exaction of structures that correspond to free energy minima out of the total conformational space. The total conformational space is generated using molecular dynamic simulations with enhanced sampling over a wide energy range by making use of a generalized (non-Boltzman) distribution functions at multiple temperatures [126]. This approach uses a biased potential for the MD simulations generated from the generalized distribution function that, when reweighted correctly, can lead to the desired thermodynamic information. This method was shown to be efficient in free energy simulation [127] as well as in protein folding studies [128].

In all simulations, Amber 99 force field [129] was used and all the quantum mechanics (QM) calculations were carried out using Gaussian 03 program suite [130]. The force field parameter set for all nucleotide fragments were generated using the RESP method described previously. Extended structure was used as initial structure for a 500-step minimization and a 20 ps heating step from 0 K to 300K. Then, the enhanced sampling simulation was conducted at 300 K with 300 β-values exponentially distributed from 250 to 1000 K to generate a much wider energy distribution. After 20 ns simulations, a set of converged weighing factors for different temperatures (i.e., different β) was obtained. Using this set of weighting factors, ten independent enhanced sampling simulations starting from initial structures were conducted for 500 ns, respectively. The

potential energy distribution of all trajectories during the iteration procedure was in a wide range as expected. With the fixed biased potentials, a largely uniform energy distribution was obtained in a wide range of ~500 kcal/mol, compared to the ~100 kcal/mol energy range sampled by the normal MD simulation. The root mean square deviation (RMSD) of DNA backbone was used as the reaction coordinate of the conformational change in DNA. The free energy vs. the RMSD distribution was computed based on the probability distribution of all conformations.

Cluster analysis

Cluster analysis is used to group the candidate structures based on their similarity. The clustering method used here was developed by MMTSB [131]. To find clusters of structures, the cartesian coordinate RMSD values between all pairs of structures were calculated. For each structure, the number of other structures for which the RMSD values less than a fixed cluster radius (2Å was used in all cluster analyses) was determined. The structure with the highest number of neighbors was taken as the center of a cluster, and formed together with all its neighbors a cluster. The structures of this cluster were eliminated from the pool of structures. The process was repeated until the pool of structures was empty. At the end, a series of non-overlapping clusters of structures was obtained. For each cluster with high population, we picked 500 to 1,000 structures that is nearest the cluster center for further collision cross-section (CCS) calculation using the trajectory method (TM) in MOBCAL software [74, 132]. The theoretical CCS profiles of each major cluster were determined using a Gaussian curve

distribution with the mean values and the standard deviation defined by the CCS calculations.

Results and Discussion

Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) of DNA

1. Choice of matrix and sample preparation for MALDI-MS of nucleic acids

It is known that the key to successfully analyze a new class of analyte compounds in MALDI-MS is the choice of matrix, sample preparation and laser wavelength. To find the most optimum experimental condition for DNA mass analysis, seven different sample preparation methods involving four matrices (3-HPA, 2, 4, 6-THAP, 2, 5-DHB, αCHCA) and twelve different length oligonucleotides (5- to 14-mer) were under investigation. Two different sample deposition methods were used: the traditional dried droplet methods and the overlayer methods. Koomen *et al.* claimed that the overlayer methods gave more homogenous crystallization and yielded more reproducible and higher quality results [133]. In addition, diammonium hydrogen citrate was used as matrix additives to suppress the adduction of alkali metal cations, *i.e.,* Na^+ and K^+ ions. Table 8 and 9 lists all of the oligonucleotides sequences studied and the molecular weights of their monomers and the sample preparation methods that are examined in this study.

All seven sample preparation methods were first tested with dGGATC using STR in positive ion mode and the representative spectra are shown in Figure 20. All methods yield results for the dGGATC sample exceptαCHCA using overlayer sample preparation
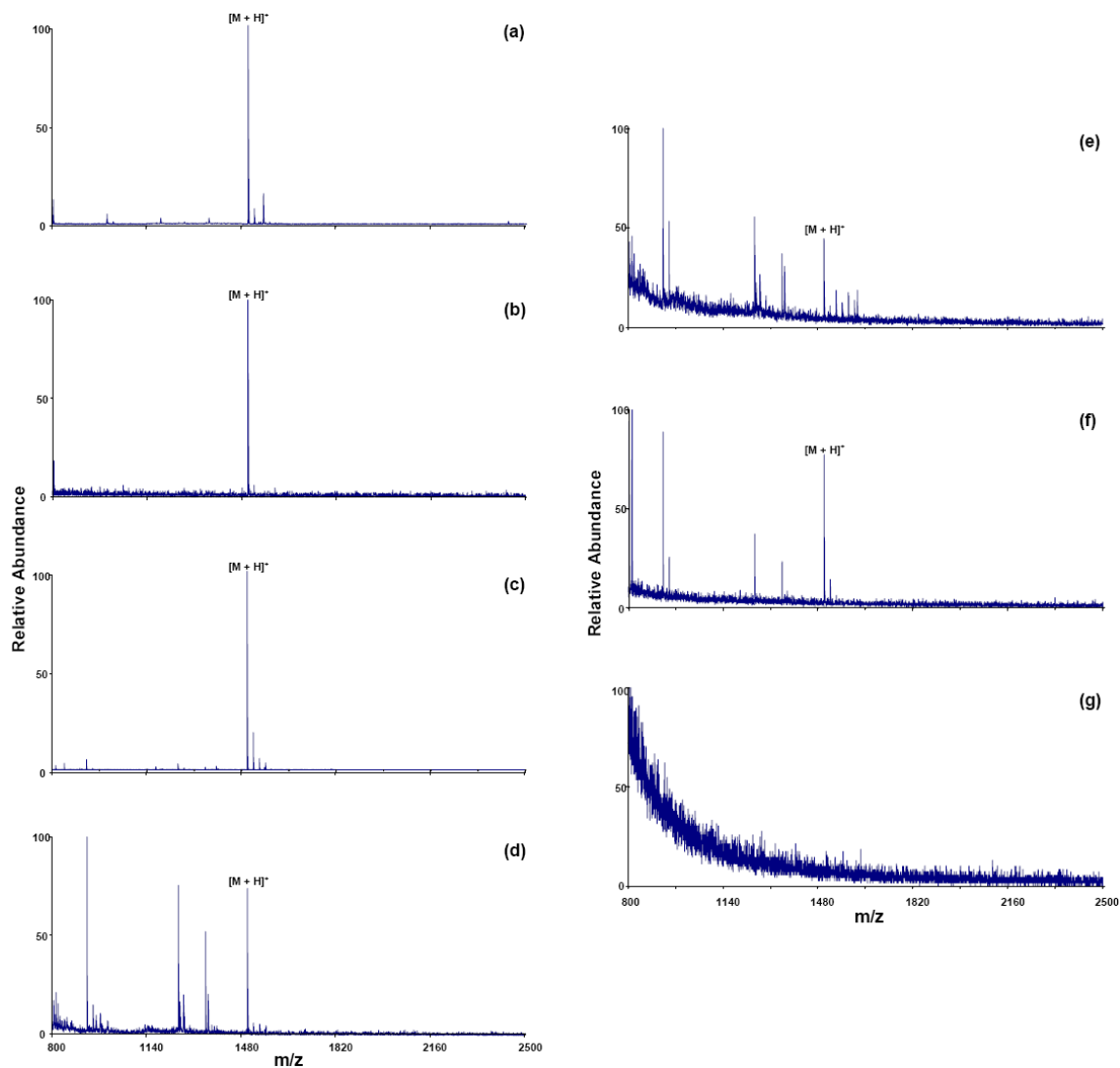
Figure 20. Positive ion mode MALDI mass spectra of dGGATC analyzed on STR with different preparation methods using N2 laser. (a) Dried droplet methods using 35mg/mL HPA matrix, (b) dried droplet methods using 20mg/mL THAP matrix, (c) dried droplet methods using 50mg/mL THAP matrix, (d) dried droplet methods using 40mg/mL DHB matrix, (e) dried droplet methods using 100mg/mL DHB matrix, (f) overlayer methods using 30mg/mL THAP matrix, (g) overlayer methods using 30mg/mLαCHCA matrix.

method. In general, all of the dried droplet sample deposits are similar in morphology and no one method was found to generate reproducible results from any location on the sample spot. The overlayer method appears to give more uniform depositions and ion yields are more reproducible both in terms of laser shot-to-shot and from sample spot-to-spot as compared to dried droplet preparations. The spectral quality of each method was evaluated in terms of signal-to-noise (S/N) ratio, resolution, mass measurement accuracy, relative abundance of alkali metal ion adducts and fragmentation of DNA analytes (Table 10). The mass spectra obtained from each method vary considerably in S/N ratio for the protonated monomer 5-mer molecules. 3-HPA deposits yield the highest ion signal (in excess of 1,300 for the protonated 5-mer ions) of all dried droplet methods. The general trend observed for S/N ratio of the protonated molecules is 3-HPA > 2, 4, 6-THAP > 2, 5-DHB. The resolution in the 40mg/mL 2, 5-DHB spectra is the highest of all dried droplet sample preparation methods: > 12,000 on the 5-mer. The resolution values observed in the other dried droplet methods are pretty close ranging from 6,000 to 7,000. The mass accuracy observed in 3-HPA is slightly better than the other matrices as well. Furthermore, this method yields minimal amounts of alkali adduct ions. Fragment ions are observed in greater abundance in spectra obtained using 2, 5-DHB as compared to other matrices we have examined using dried droplet preparations. The overlayer sample preparation methods do not appear to give better mass spectral data than dried droplet methods as expected. The overlayer deposits of αCHCA do not produce any signal for the 5-mer molecules.

Table 10. The spectral quality of each method acquired by STR for dGGATC$^+$ ions was evaluated in terms of signal-to-noise (S/N) ratio, resolution, mass measurement accuracy, relative abundance of alkali metal ion adducts and fragmentation of DNA analytes.

Method	No	S/N	Resolut ion	Mass accuracy (ppm)	Relative abudance of Na adducts ion from ionized molecule	Fragmentation
Dried droplet	1	1365	5945	63	7.8%	little
	2	232	6696	96	6.0%	no
	3	650.8	5474	309	20.5%	lttle
	4	237	12796	744	8.4%	lot
	5	41.3	6707	697	24.3%	lot
Overlayer	6	103.4	7675	166	18.6%	lot
	7	0	0	0	0.0%	

Spectrum acquired from sample made with αCHCA and 2, 4, 6-THAP showed significant loss from the signal of the $[M + H]^+$ ions of dGGATC due to fragmentation. The use of 3-HPA with traditional dried droplet method seems to be the best sample preparation method for the DNA analysis performed with STR.

We also examined all sample preparation methods with dGGATC using 4700 in both positive and negative ion modes. 4700 uses Nd:YAG laser ($\lambda = 355$ nm) in stead of N_2 laser ($\lambda = 337$ nm) used by STR. Mass spectra acquired by 4700 in both ion modes are shown in Figures 21 and 22 and the values for evaluating the mass spectra obtained from the different sample preparation methods are listed in Table 11. Significant improvement in both resolution and mass accuracy is observed using 4700 when compared to STR. Among all positive ion mass spectra, only three dried droplet methods gave good results including 50 mg/mL 2, 4, 6-THAP with ammonium matrix additive, 40 mg/mL 2, 5-DHB with ammonium matrix additive and 100 mg/mL 2, 5-DHB in MeOH. The ion yields for 2, 5-DHB deposits are generally higher than 2, 4, 6-THAP, however, more fragmentation ions are observed in spectra. In negative ion mode, all methods were found to yield more reproducible results except sample made with 3-HPA. The overlayer sample preparation methods appeared to give very high S/N ratio, but fragmentation of DNA analytes is a mjor problem when compared to dried droplet preparations. In general, deposit of 2, 5-DHB at 40 mg/mL with ammonium matrix additive so far gave the best mass spectral data in both positive and negative ion modes when using 4700 for DNA mass analysis.
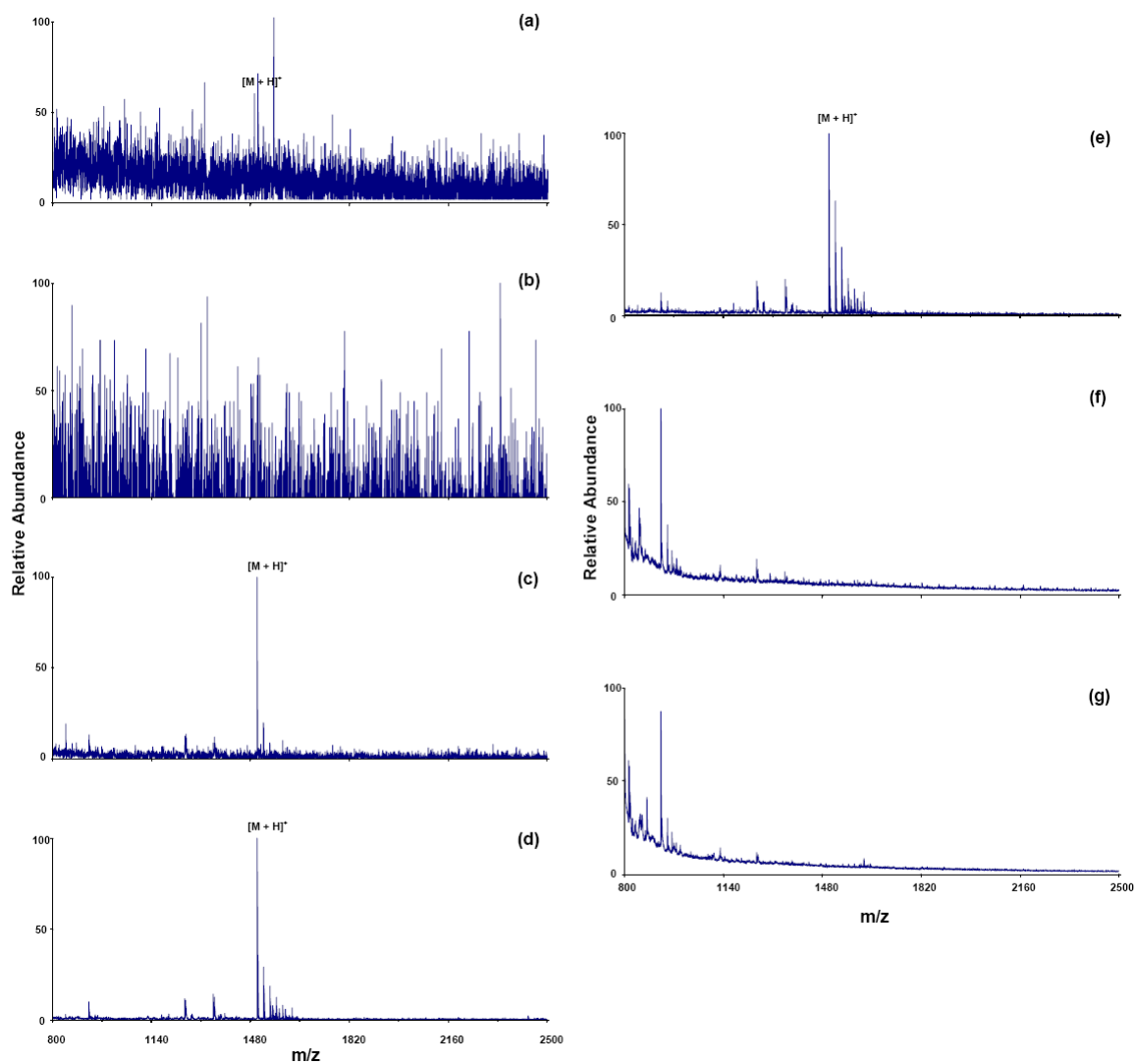
Figure 21. Positive ion mode MALDI mass spectra of dGGATC analyzed on 4700 with different preparation methods using YAG laser. (a) Dried droplet methods using 35mg/mL HPA matrix, (b) dried droplet methods using 20mg/mL THAP matrix, (c) dried droplet methods using 50mg/mL THAP matrix, (d) dried droplet methods using 40mg/mL DHB matrix, (e) dried droplet methods using 100mg/mL DHB matrix, (f) overlayer methods using 30mg/mL THAP matrix, (g) overlayer methods using 30mg/mLαCHCA matrix.
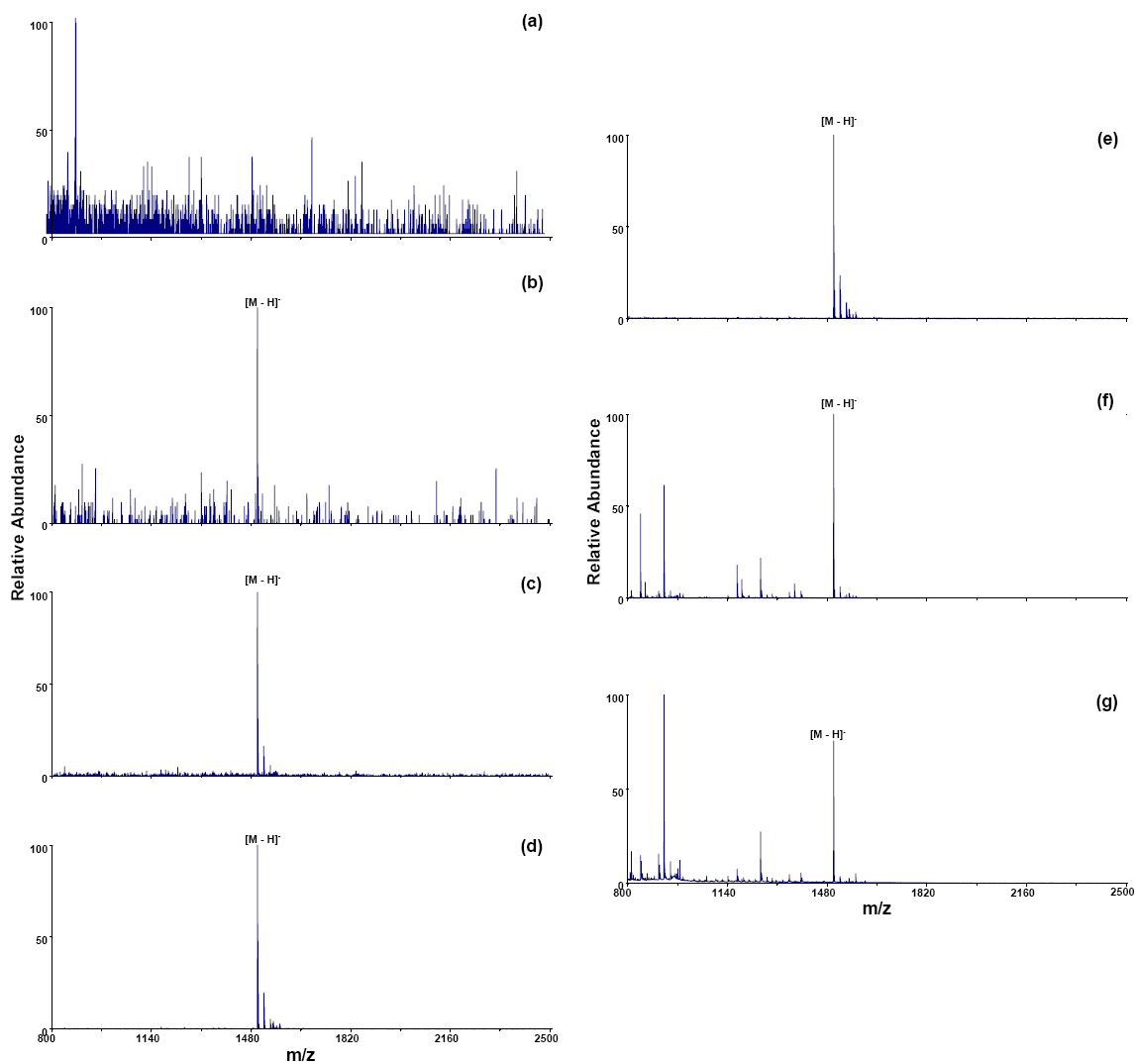
Figure 22. Negative ion mode MALDI mass spectra of dGGATC analyzed on 4700 with different preparation methods using YAG laser. (a) Dried droplet methods using 35mg/mL HPA matrix, (b) dried droplet methods using 20mg/mL THAP matrix, (c) dried droplet methods using 50mg/mL THAP matrix, (d) dried droplet methods using 40mg/mL DHB matrix, (e) dried droplet methods using 100mg/mL DHB matrix, (f) overlayer methods using 30mg/mL THAP matrix, (g) overlayer methods using 30mg/mLαCHCA matrix.

Table 11. The spectral quality of each method acquired by 4700 for dGGATC ions in both ion modes was evaluated in terms of signal-to-noise (S/N) ratio, resolution, mass measurement accuracy, relative abundance of alkali metal ion adducts and fragmentation of DNA analytes.

Positive ion mode

Method	No	S/N	Resolution	Mass accuracy (ppm)	Relative abudance of Na adducts ion from ionized molecule	Fragmentation
Dried droplet	1	12.2	20372	13	0.0%	no
	2	0	0	0	0.0%	no
	3	113.6	16385	21	19.7%	not much
	4	557.5	16487	14	29.3%	lot
	5	341	17412	25	62.9%	worst
Overlayer	6	9.7	10875	4	0.0%	lot
	7	0	0	0	0.0%	

Negative ion mode

Method	No	S/N	Resolution	Mass accuracy (ppm)	Relative abudance of Na adducts ion from ionized molecule	Fragmentation
Dried droplet	1	0	0	0	0.0%	no
	2	10	27068	49	0.0%	no
	3	120.8	16543	79	16.3%	no
	4	1543.9	19064	10	19.6%	little
	5	787.9	18324	4	23.4%	little
Overlayer	6	7318.2	13068	3	6.3%	lot
	7	1557.3	10908	18	4.7%	lot

The next series of experiments was done on different length oligonucleotides using both 3-HPA and 2, 5-DHB as matrix acquired by STR and 4700 with both negative and positive ion modes, respectively. Representative mass spectra are shown in Figures 23 – 30, where studies with 3-HPA using different instrument in different polarities are compared in Figures 23 – 24, 27 - 28 and studies with 2, 5-DHB are compared in Figures 25 – 26, 29 – 30. The S/N ratios and mass resolutions of all nucleotides observed in above spectra studied by STR and 4700 are represented in Figures 31 and 32, respectively. As can be seen, the abundance of the protonated or deprotonated analyte molecules decreased dramatically with the increase of the nucleotide size. No peaks of the protonated or deprotonated analyte ions were detected for the analysis of oligonucleotides up to 12-mer under any conditions we have examined with the use of 2, 5-DHB. 3-HPA seems to work better than 2, 5-DHB for the analysis of large oligonucleotides when using STR. The S/N ratio ranges from around 3,000 for the protonated 6-mer to *c.a.* 30 for the protonated 14-mer using STR and the signal for the deprotonated nucleotides are in the range 70 – 4,000 (Figure 31). To our delight, the mass resolution doesn't decrease much as the mass of the analyte ions increases in all the cases we have examined. Only a little bit peak-tailing was observed for the higher mass analytes. As expected, 3-HPA as matrix yields a better signal-to-noise ratio and less fragmentation than 2, 5-DHB when using STR, while 2, 5-DHB works better for those studies performed with 4700. Furthermore, oligonucleotides are more readily being observed in positive ion mode when using 4700 and STR yields a stronger signal and better mass resolution for the detection of negative nucleotides ions than of positive ions.
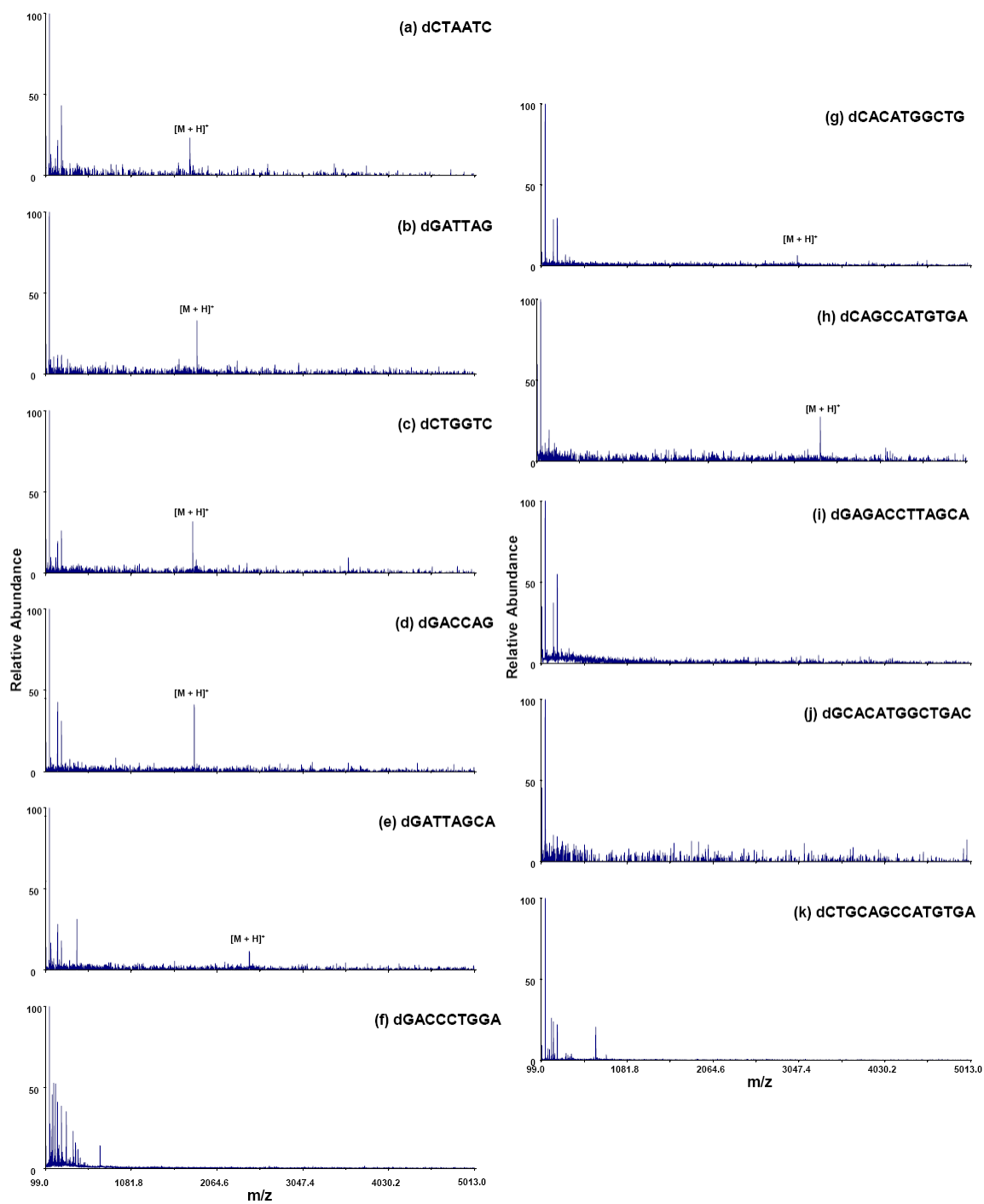
Figure 23. Positive ion mode MALDI mass spectra of all oligonucleotides analyzed by 4700 using 35 mg/mL 3-HPA as matrix and YAG laser.
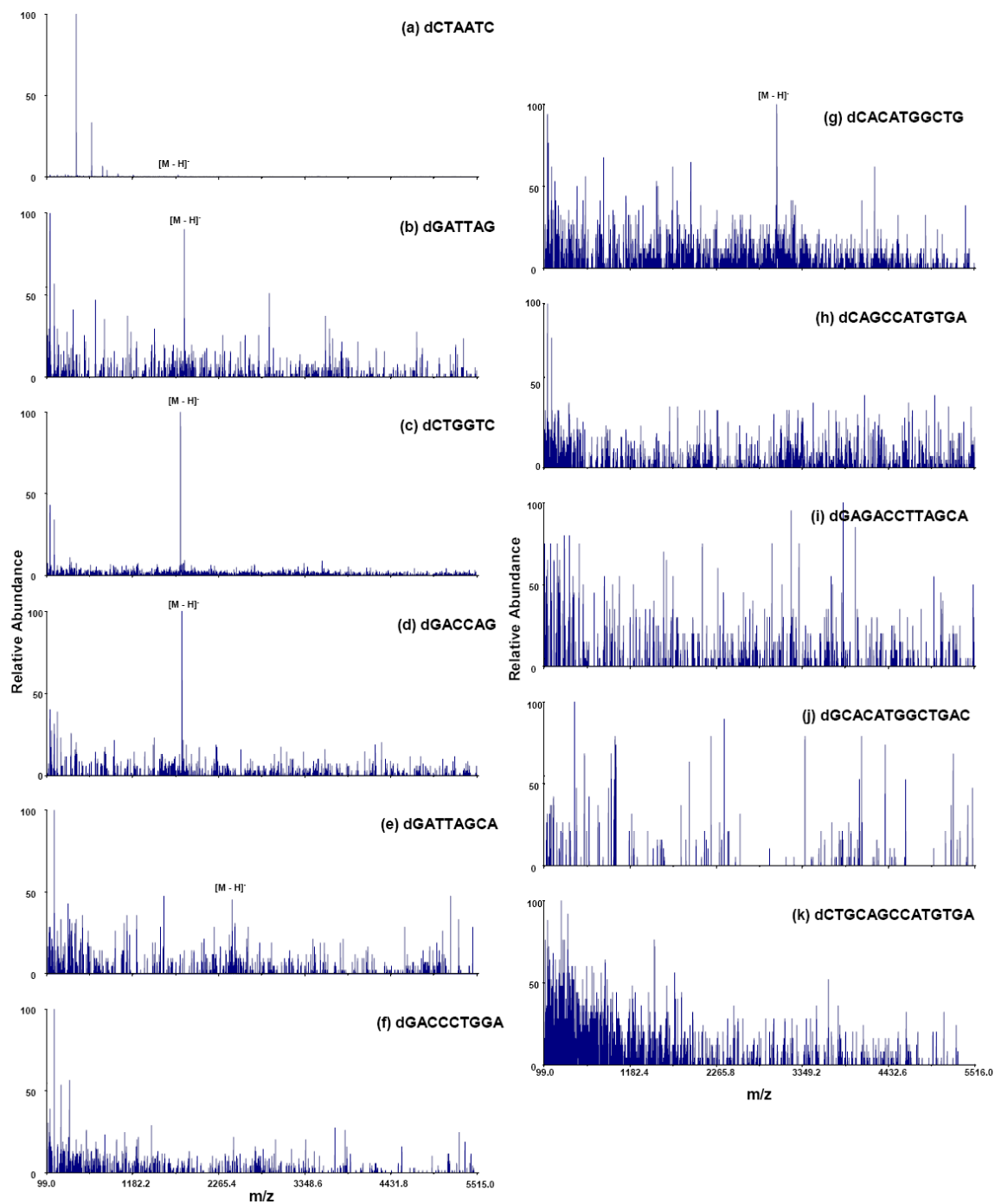
Figure 24. Negative ion mode MALDI mass spectra of all oligonucleotides analyzed by 4700 using 35 mg/mL 3-HPA as matrix and YAG laser.
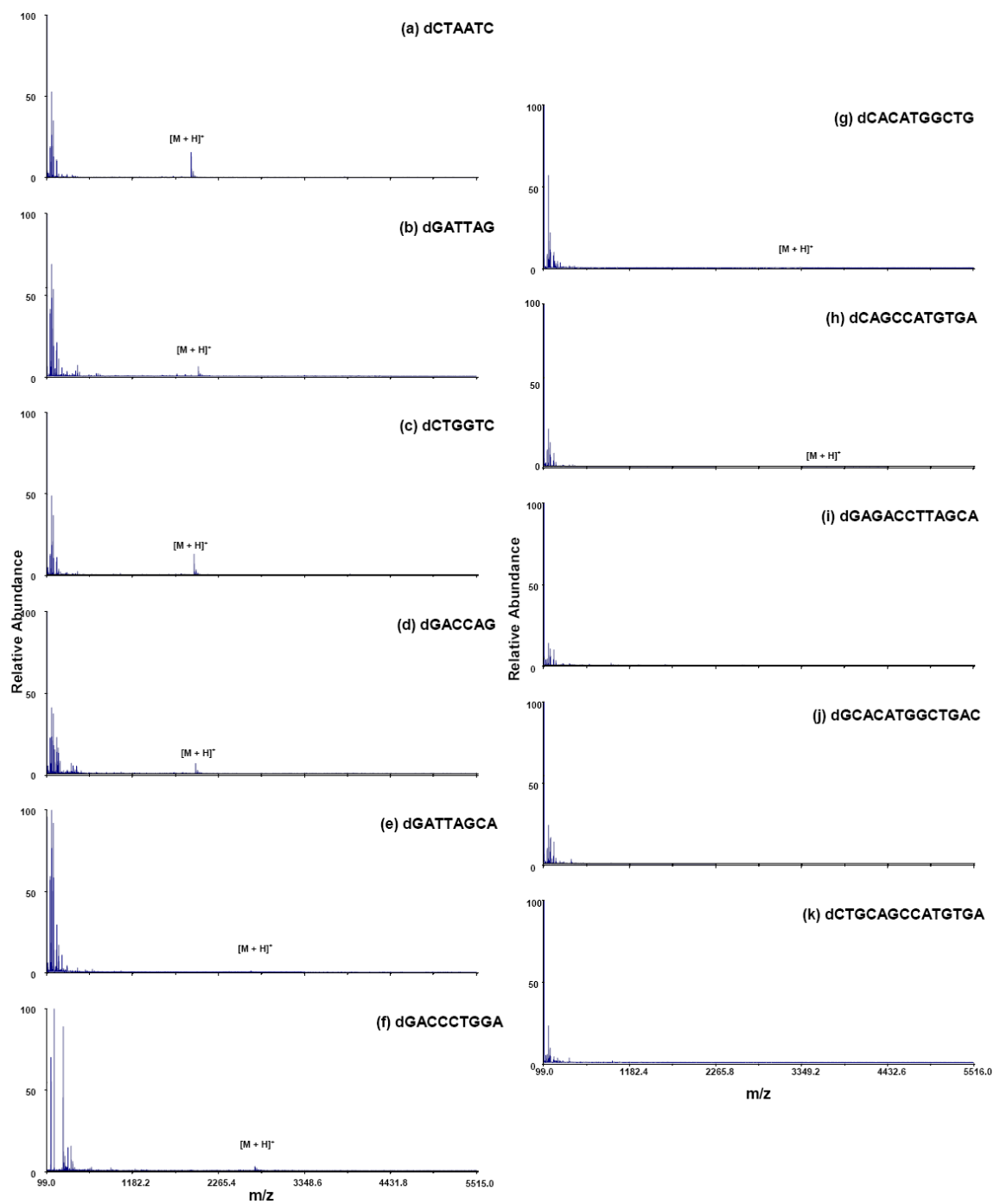
Figure 25. Positive ion mode MALDI mass spectra of all oligonucleotides analyzed by 4700 using 40 mg/mL 2, 5-DHB as matrix and YAG laser.
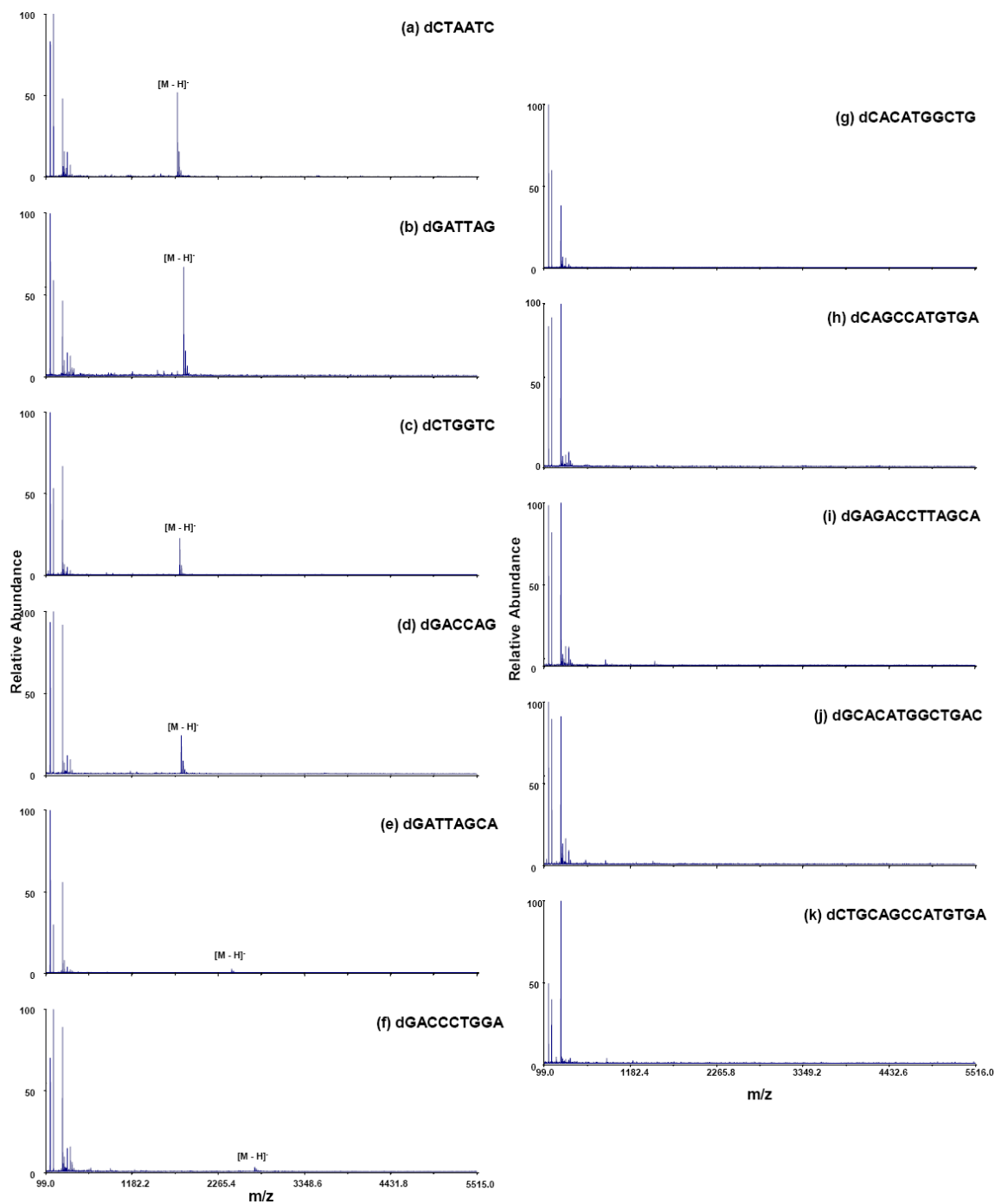
Figure 26. Negative ion mode MALDI mass spectra of all oligonucleotides analyzed by 4700 using 40 mg/mL 2, 5-DHB as matrix and YAG laser.
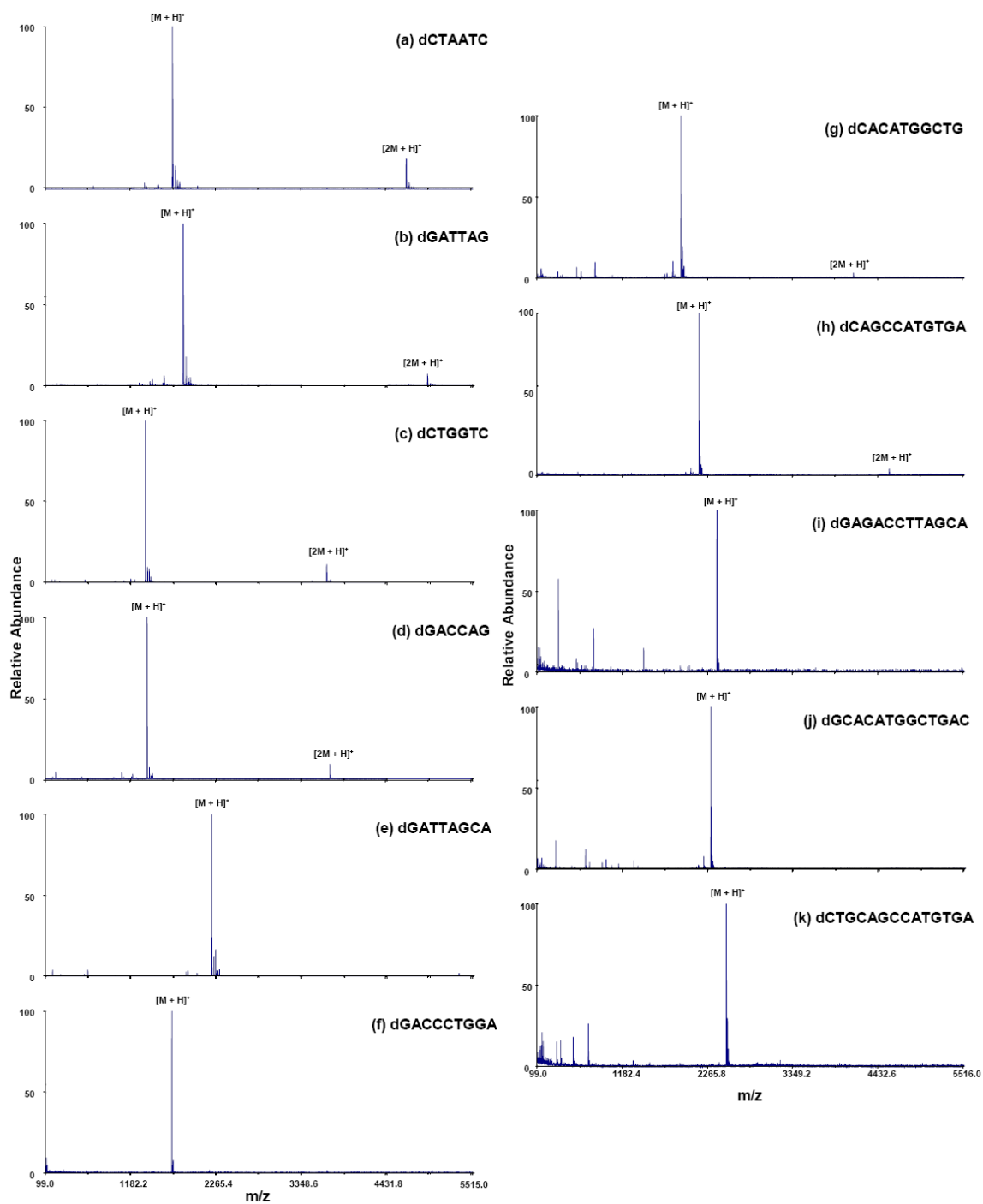
Figure 27. Positive ion mode MALDI mass spectra of all oligonucleotides analyzed by STR using 35 mg/mL 3-HPA as matrix and N_2 laser.
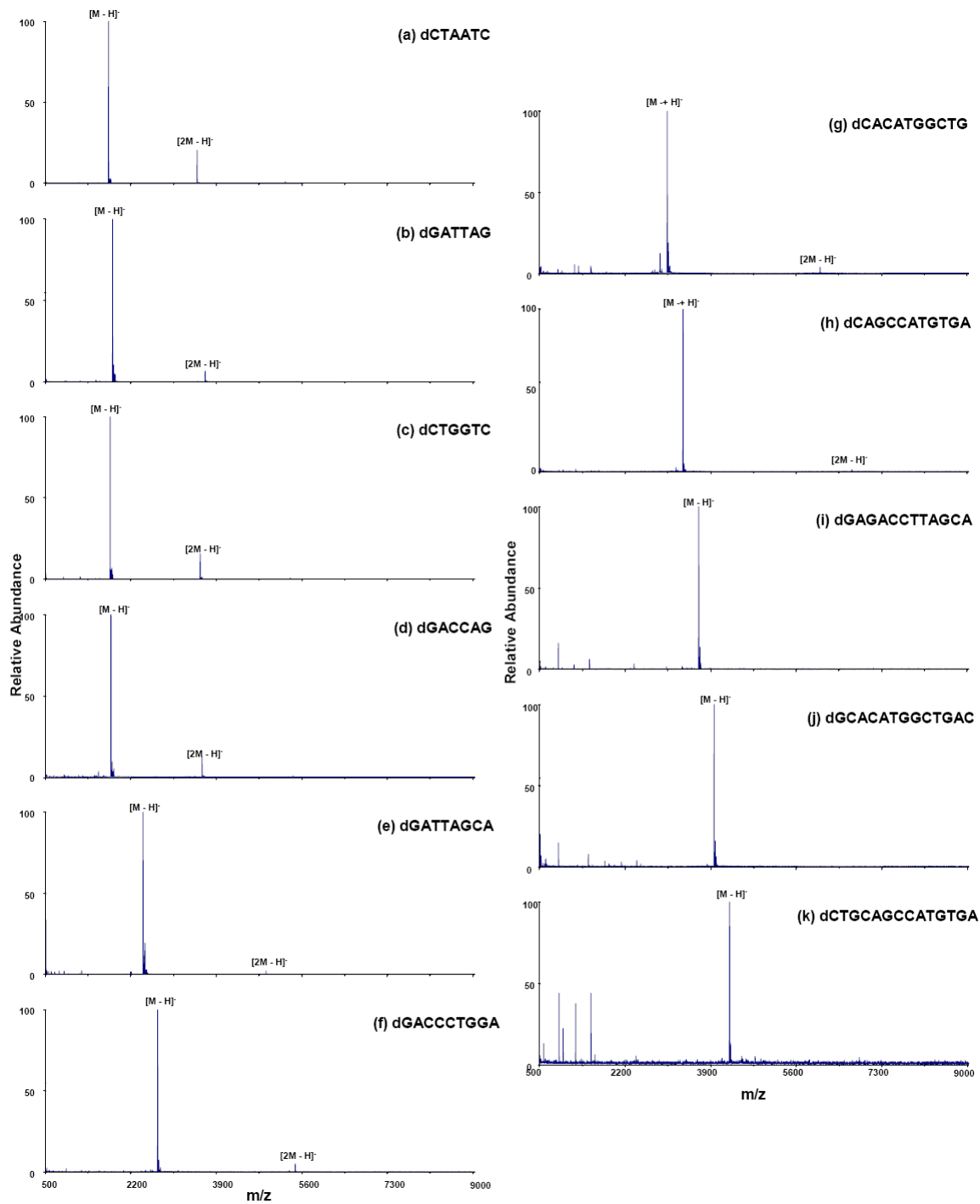
Figure 28. Negative ion mode MALDI mass spectra of all oligonucleotides analyzed by STR using 35 mg/mL 3-HPA as matrix and N_2 laser.
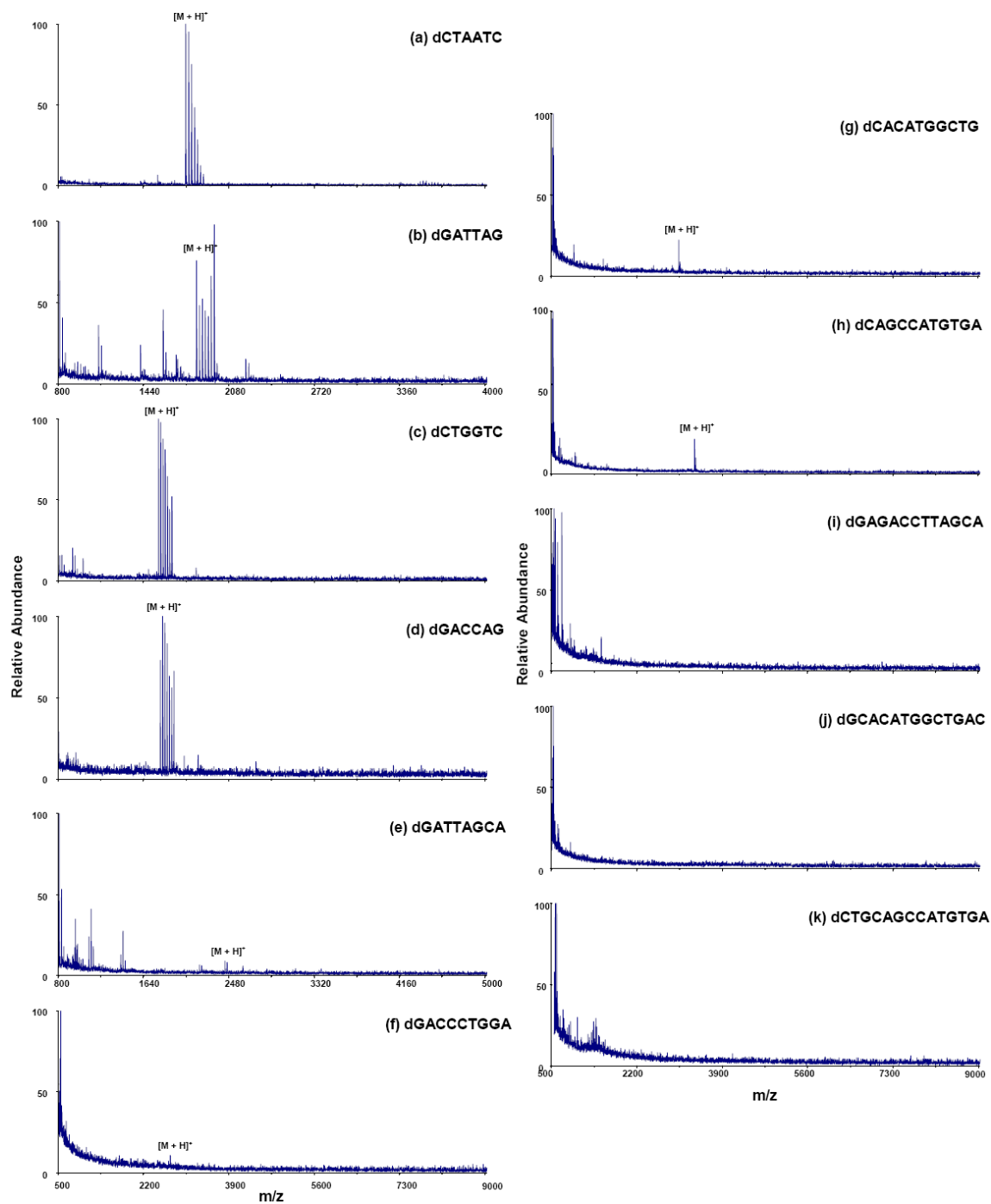
Figure 29. Positive ion mode MALDI mass spectra of all oligonucleotides analyzed by STR using 40 mg/mL 2, 5-DHB as matrix and N_2 laser.
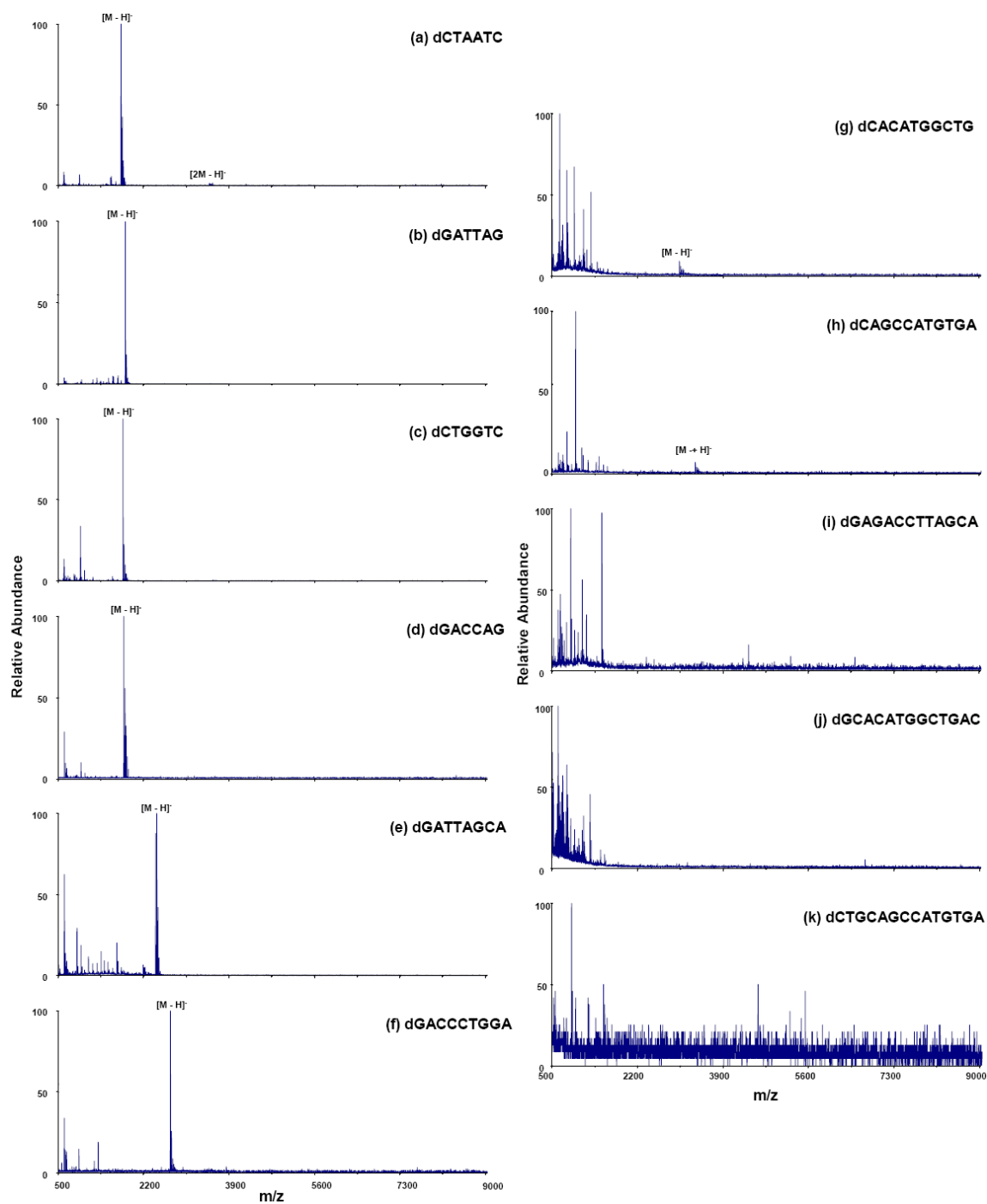
Figure 30. Negative ion mode MALDI mass spectra of all oligonucleotides analyzed by STR using 40 mg/mL 2, 5-DHB as matrix and N_2 laser.

a)

S/N-STR

b)
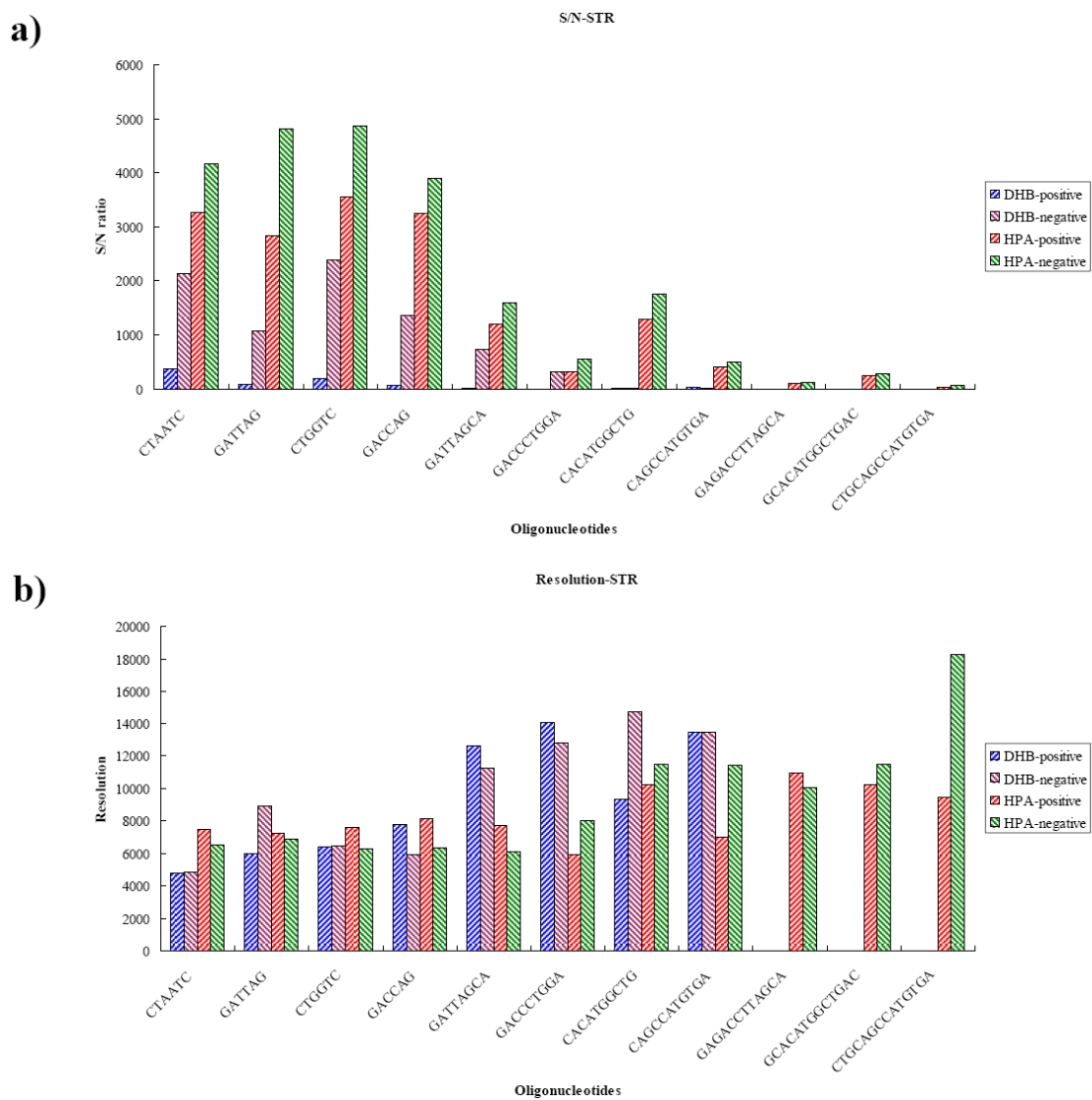
Resolution-STR

Figure 31. Average S/N ratios and resolutions of all oligonucleotides observed by STR using different matrices in different ion modes.

a)

b)
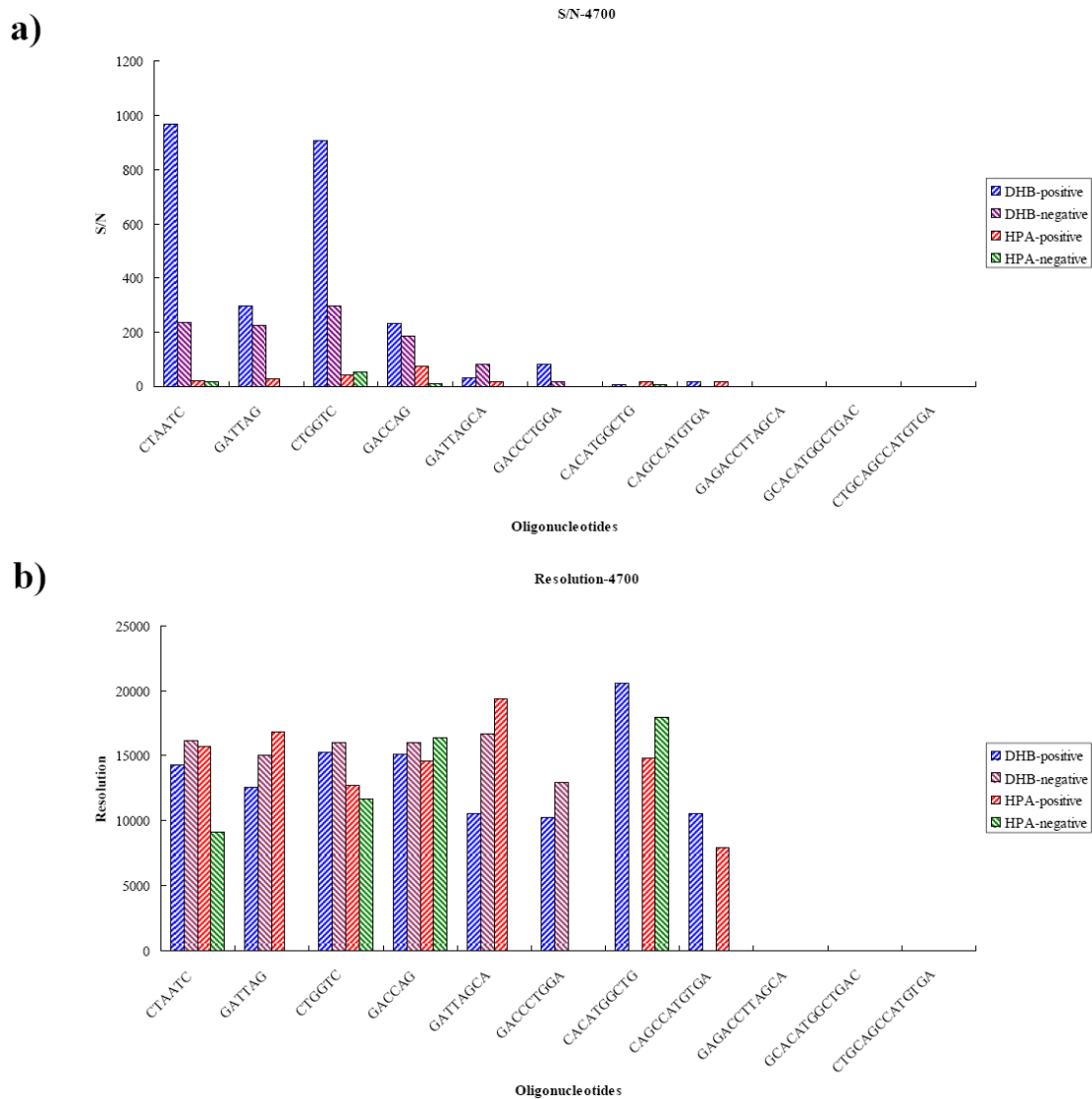
Figure 32. Average S/N ratios and resolutions of all oligonucleotides observed by 4700 using different matrices in different ion modes.

2. Laser test for MALDI-MS of nucleic acids

As different instruments (STR and 4700) were found to give quite different spectral profiles for the same sample spot, the influence of different laser wavelength on MALDI-MS of nucleotides was investigated next using frequency-tripled Nd:YAG laser (wavelength of 355 nm) on STR. All nucleotides were tested with 3-HPA and 2, 5-DHB in both polarities, respectively and representative mass spectra are shown in Figures 33 - 36. The comparisons of S/N ratios of the oligonucleotides with 3-HPA and 2, 5-DHB using different laser wavelength are illustrated in Figure 37. With YAG laser, STR yields a pretty similar spectral profile for the nucleotides as with Nitrogen laser, albeit at a lower abundance. In general, nucleotides with 3-HPA give stronger signals than with 2, 5-DHB and negative ions are easier detected than positive ions which is opposite of the trend observed on 4700. It is thus clear that the different results observed between STR and 4700 are not a consequence of using different laser.

3. Study of non-covalent complexes for MALDI-MS of nucleic acids

When oligonucleotides were tested with 3-HPA on STR, it was found that the singly charged monomer of the oligonucleotide is always the dominant ion observed in the mass spectrum despite of the polarity. However, most interesting is the appearance of a group of ions at twice the mass of the monomer species, which is confirmed as the noncovalent dimer of the oligonucleotides. Such noncovalent dimers are observed in every oligonucleotides we have examined if the length of the oligomer is shorter than 12-mer and/or the concentration of nucleotide is greater than or equal to 300 μM.
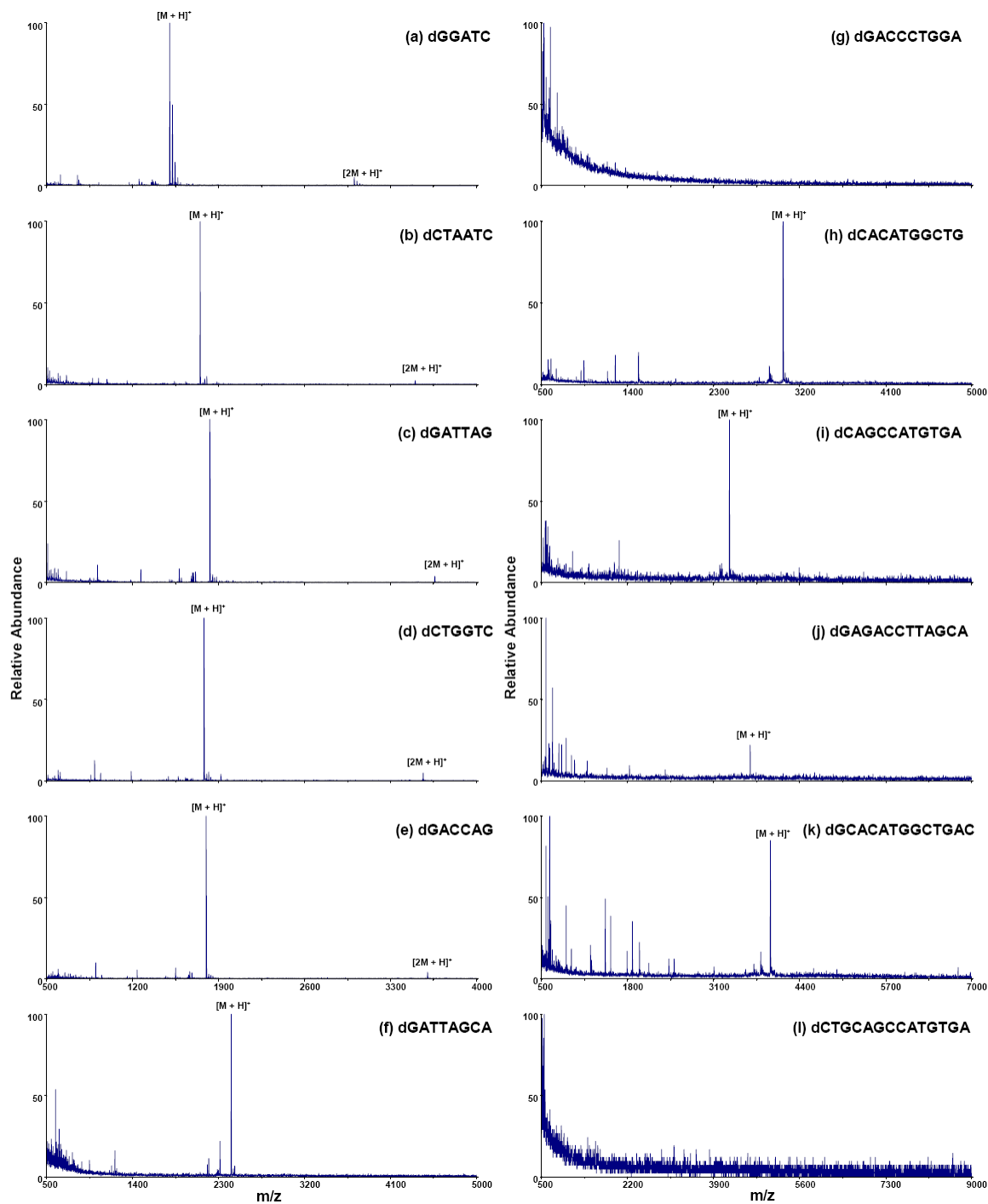
Figure 33. Positive ion mode MALDI mass spectra of all oligonucleotides analyzed by STR using 35 mg/mL 3-HPA as matrix and YAG laser.
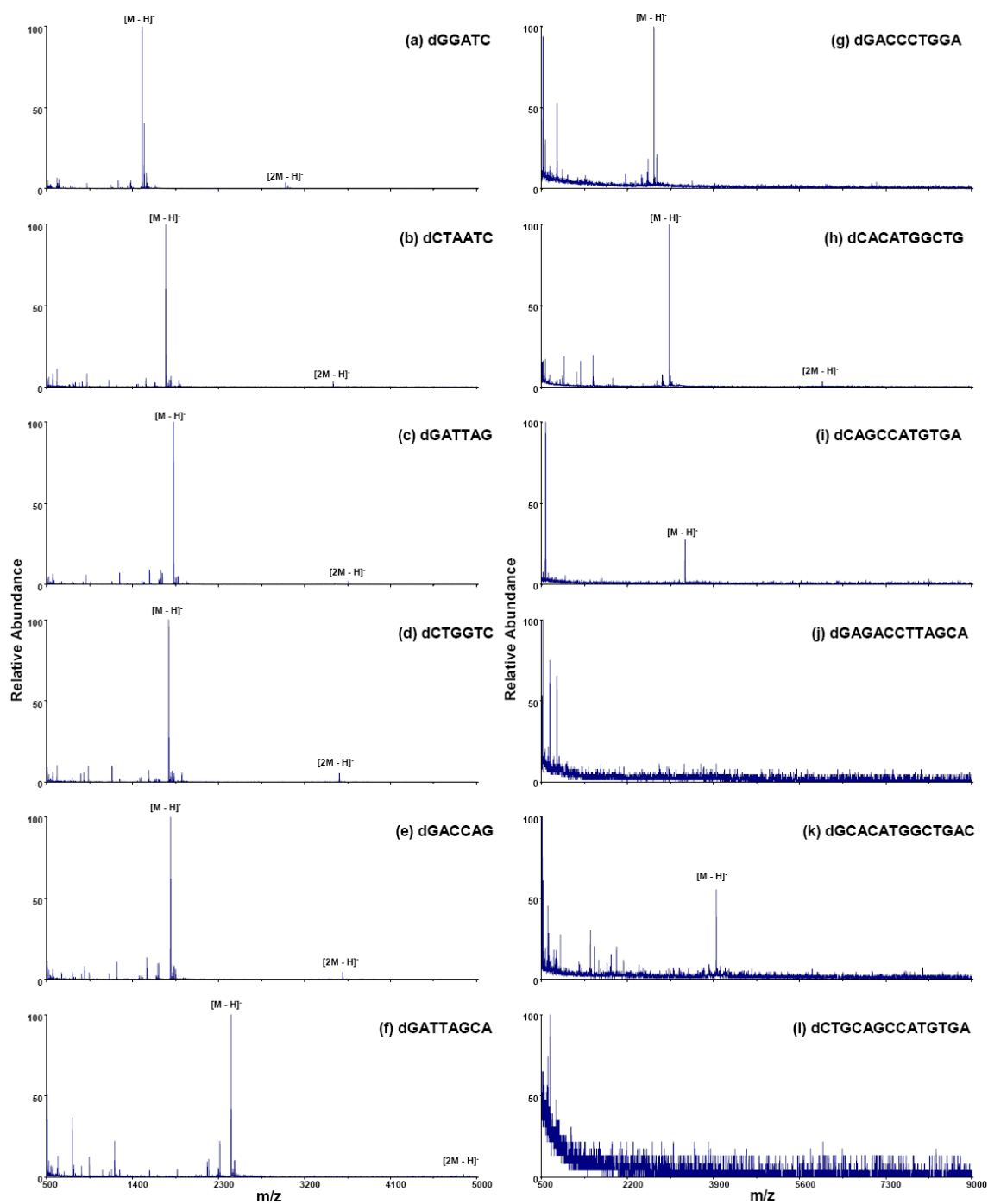
Figure 34. Negative ion mode MALDI mass spectra of all oligonucleotides analyzed by STR using 35 mg/mL 3-HPA as matrix and YAG laser.
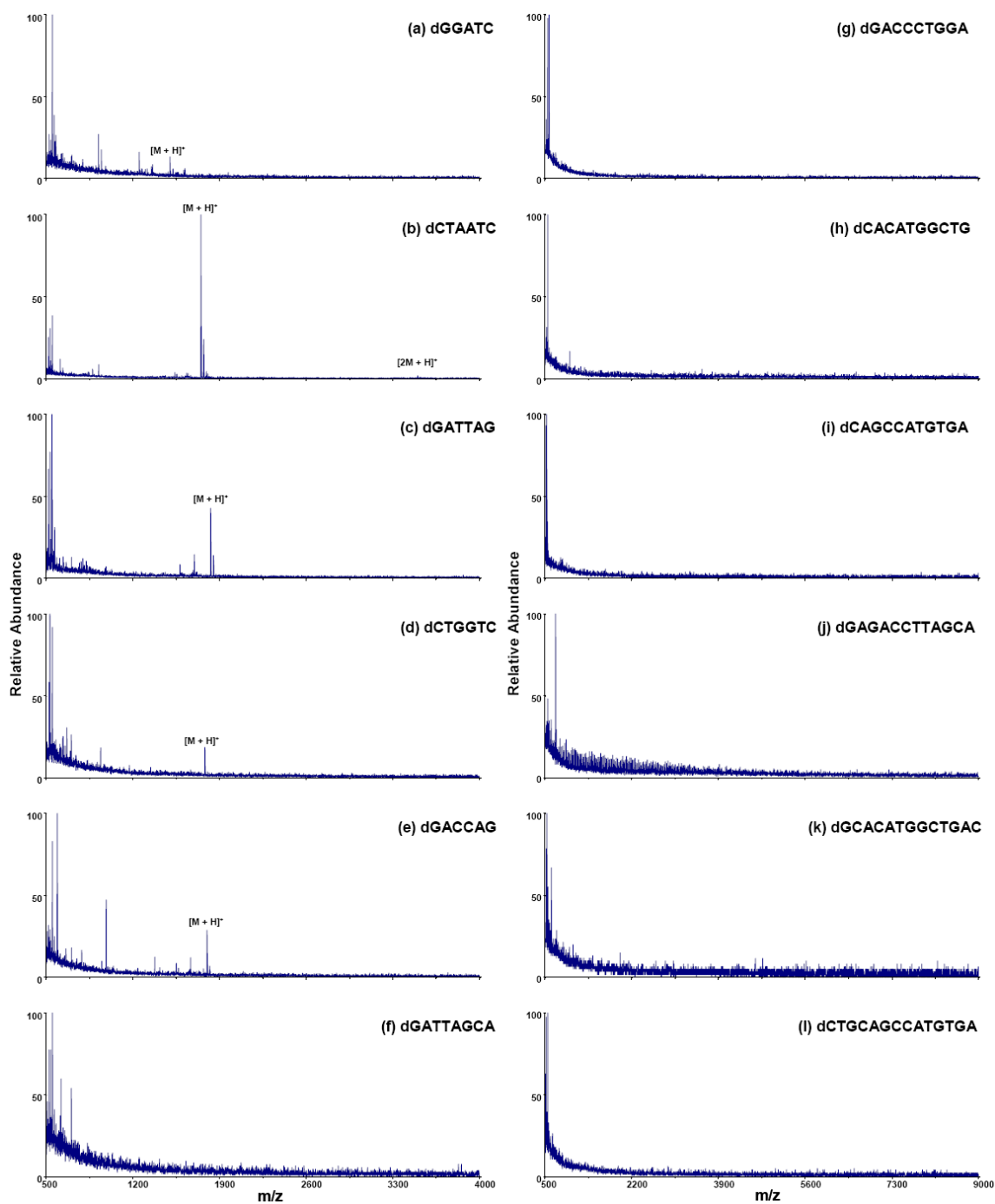
Figure 35. Positive ion mode MALDI mass spectra of all oligonucleotides analyzed by STR using 40 mg/mL 2, 5-DHB as matrix and YAG laser.
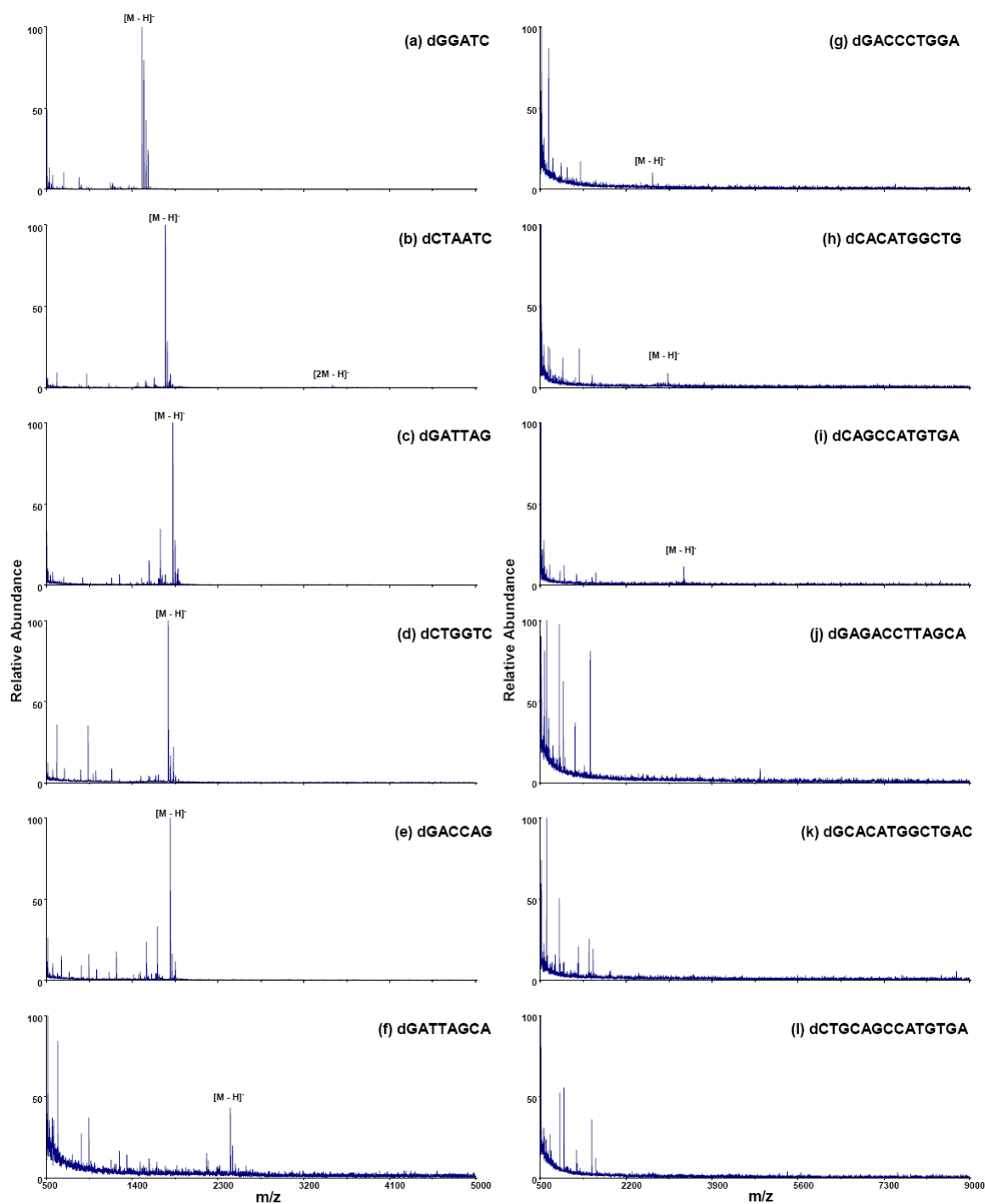
Figure 36. Negative ion mode MALDI mass spectra of all oligonucleotides analyzed by STR using 40 mg/mL 2, 5-DHB as matrix and YAG laser.
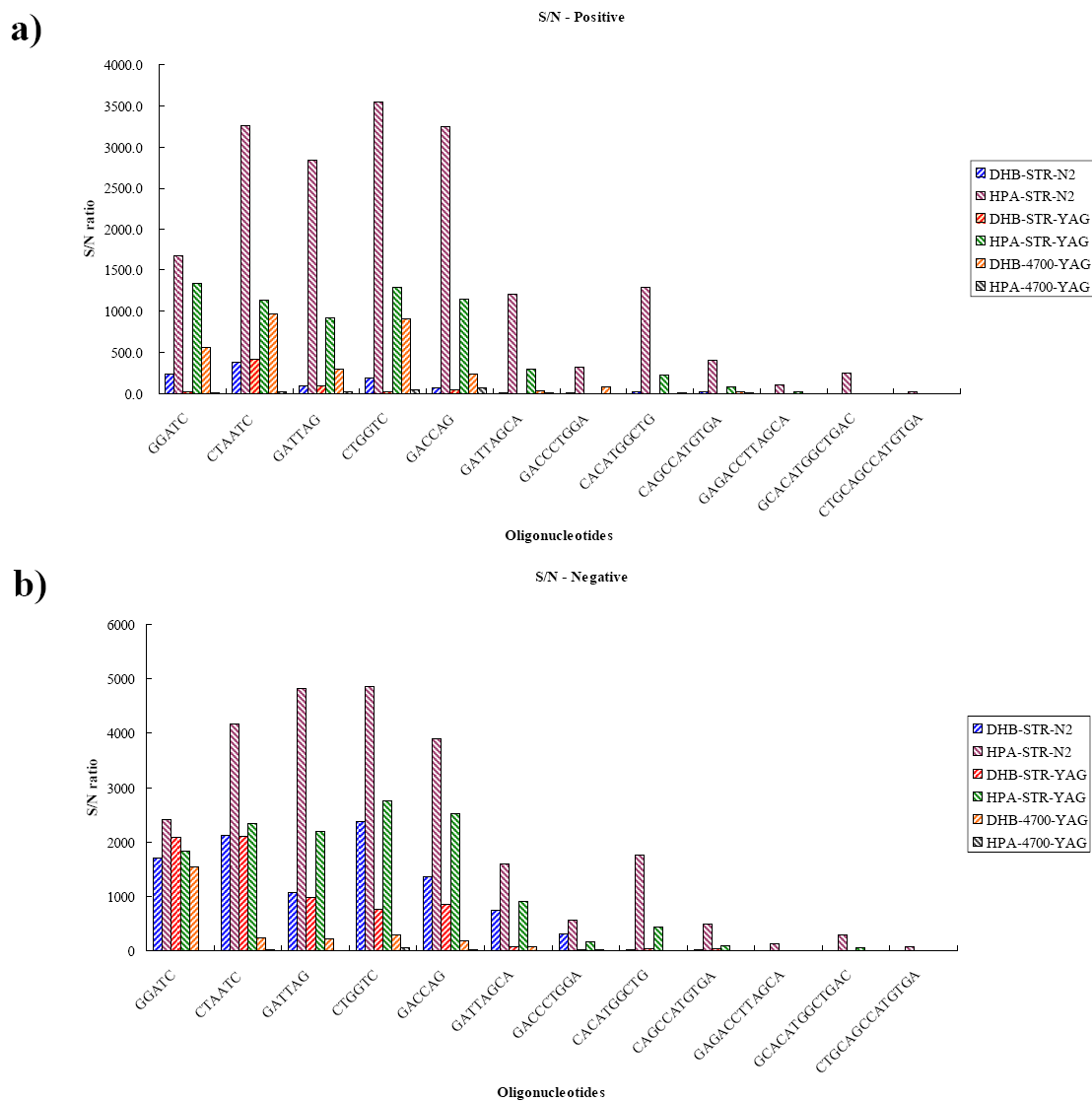
Figure 37. Average S/N ratios and resolutionse of all oligonucleotides observed by 4700 using different matrices and different wavelength lasers in different ion modes.

Interestingly, the sequence of the nucleotides in this study is not self complementrary, for example, 5-mer dGGATC, only four of the five bases can form Watson-Crick base pairs between two strands of dGGATC. A similar observation of such dimers in ESI mass spectra was made by Ding *et. al.*[134], dimers were observed even for dT_6 at a concentration of 100 μM where no self-complementary Waston-Crick pairing is possible.

To evaluate the influence of varying the relative molar ratio between analyte and matrix on the detection of noncovalent dimers of the oligonucleotides, eight different molar ratios between analyte 5-mer dGGATC and matrix 3-HPA were used. By varying the concentration of 3-HPA solution while keeping analyte dGGATC concentration constant at 500 μM, the positive ion mass spectra of each case were obtained by using STR. All measurements were repeated for five times over a week and the average abundance ratios between monomer and dimer ions observed for each case are summarized in Figure 38. No significant differences are observed in the relative abundance of the noncovalent dimer from each case. However, more sodium adducts could be seen along with an increase in the molar ratio between analyte and matrix. It seems that the molar ratio of 1:400 gives overall the best results in terms of S/N ratios, resolution, and reproducibility of the spectra when compared to other molar ratios. Thus, we use this molar ratio between analyte and matrix in the following noncovalent complex experiments.
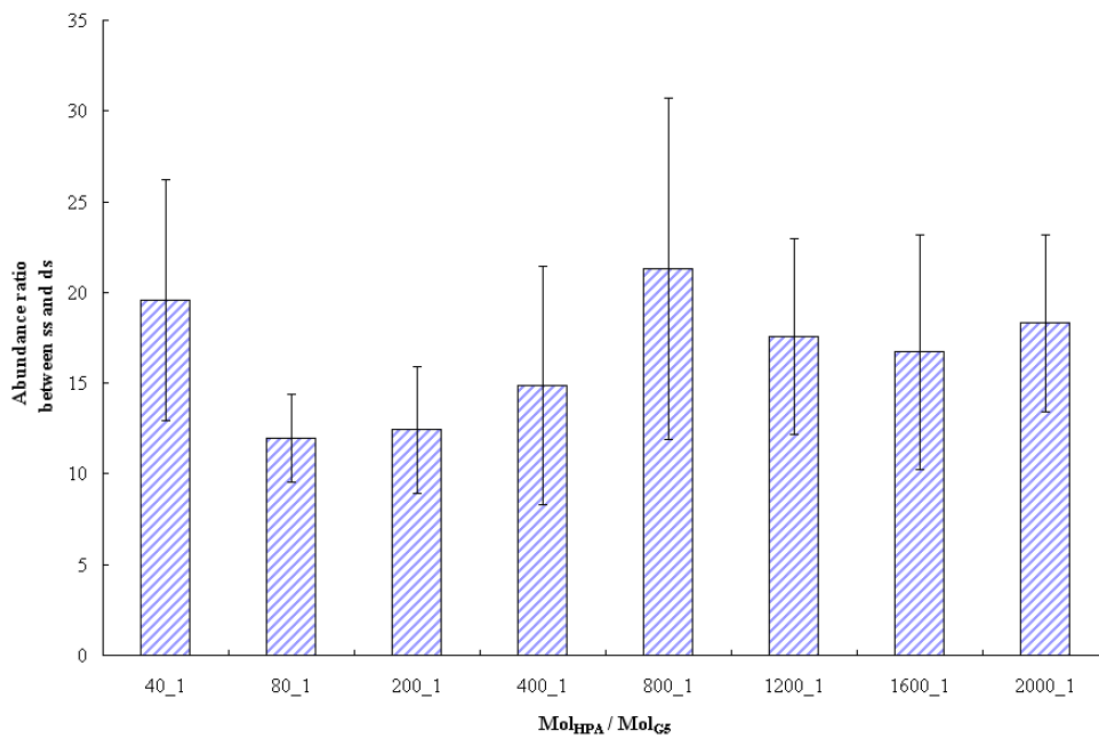
Figure 38. Average abundance ratios between monomer and dimer ions observed by STR using eight different molar ratios between matrix 3-HPA and analyte dGGATC.

To study the effect of the base sequence on the production of the noncovalent complex ions, the mixture of different oligonucleotide strands were used. Now dimmers can form between the same strands (homodimers) or between the different strands (heterodimers). The relative abundances of the homo- and heterodimers can be used to determine if there is specific interaction between two strands or not [134]. If the mixture is a nonspecific formation (i.e., the interaction is purely statistical), the predicted ratio of the homodimer:heterodimer:monomerdimer will be 25:50:25 for the monomers were equal in concentrations and had the same ionization efficiencies. If there is an excess in the relative abundance of the heterodimer over the predicted ratio, it indicates some degree of specificity in the interaction of heterodimer that does not exist for the homdimers. Hexamer dGACCAG were mixed with three different hexamers at same concentrations, respectively. The latter hexanucleotides include dCTGGTC, the complementary strands of dGACCAG, and two noncomplementary sequences, dCTAATC and dGACCAG. Three mixtures contain different numbers of Watson-Crick base pairing. The mass spectra of the mixtures are shown for comparison in Figure 39a-c. The monomers, homodimers, and heterodimers were all observed in both the mixtures of complementary and noncomplementary oligonucleotides. When closely examined the spectra, the observed ratio of the homodimer:heterodimer:monomerdimer is 25:41:34 for the complementary mixtures, 8:26:65 for the noncomplementary mixtures of dGACCAG and dCTAATC, and 26:48:26 for the mixtures of dGACCAG and dGATTAG.
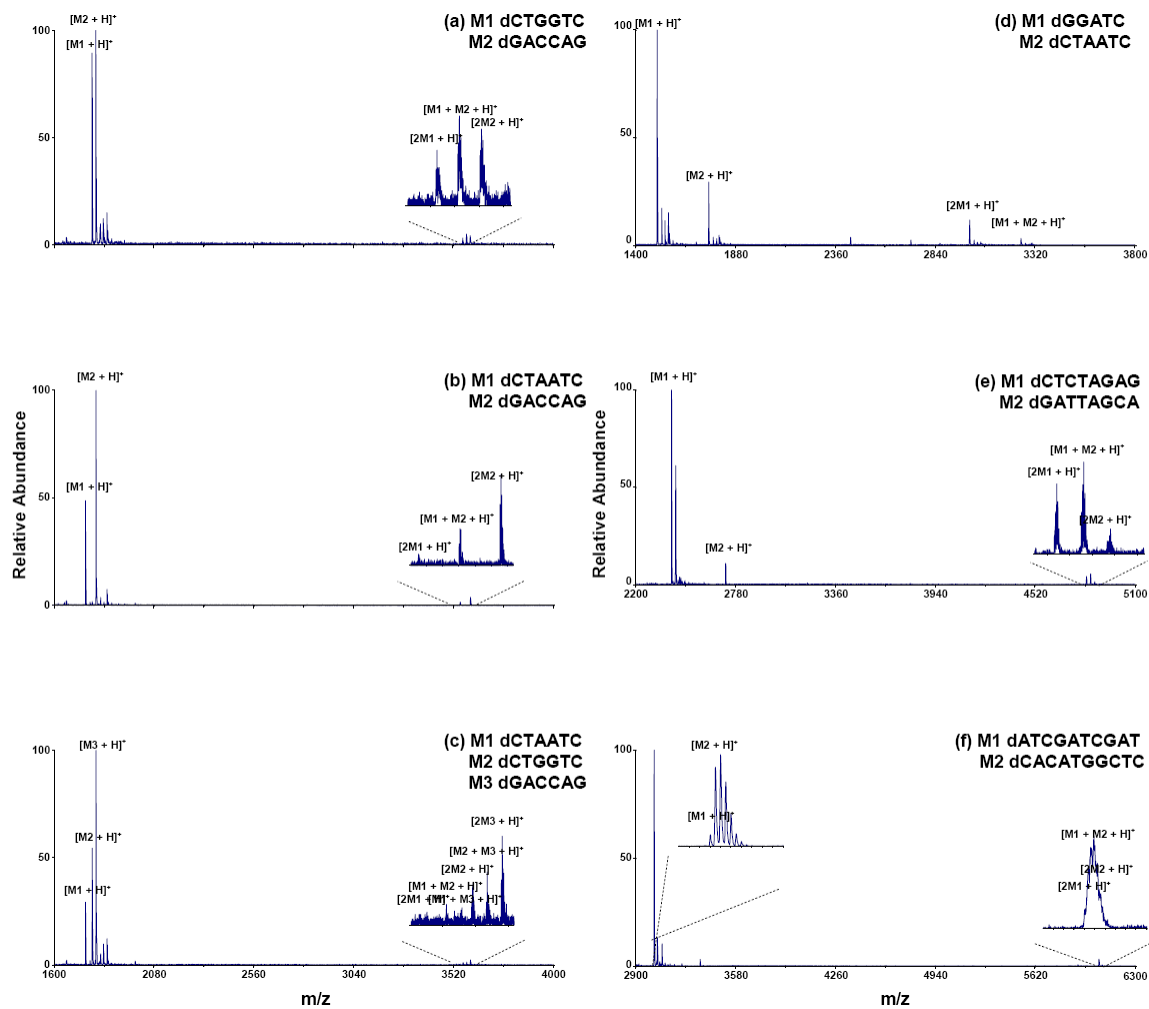
Figure 39. Positive ion mode MALDI mass spectra of oligonucleotides mixtures analyzed by STR using 35 mg/mL 3-HPA as matrix and N_2 laser.

Although the abundance distributions of the monodimers and the heterodimer are quite different among three cases, the ions from the heterodimer in all mixtures are less abundant that might be predicted on the basis of statistics, which suggests there is no specificity in the interaction. It is not surprise that there are no specific duplexes found in hexamer mixtures, as the predicted melting temperature (T_m) for the hexamer should be below room temperature, thus at the temperature of our experiment no stable duplex would be expected in the sample solution.

To observe specific duplexes, the mixtures of longer oligonucleotides were used. We explored octamers and decamers by using a pair of noncomplementary sequences in both cases. The mass spectrum of octamers (400 μM per strand) is shown in Figure 39d. As with the hexamers, the spectrum shows the presence of both the homodimers and the heterodimers and gives a ratio of 38:49:13. However, the relative abundance of the heterodimer is still under the predicted value, which indicates no specific duplexes are observed for octamer mixtures. Figure 39d shows the mass spectrum of the mixture of decamers (300 μM per strand): dATCGATCGAT and dCACATGGCTC. Since the two strands only have one unit mass difference, we are not able to resolve the peaks for monodimers and/or heterodimer and predict a ratio among them. All the results presented above for mixtures suggest that no specific duplexes exist in the mixtures we have tested. In deed, such non specific aggregates are often observed in the spectra of MALDI-MS. The signal intensity of such non specific aggregates usually increases with increasing analyte concentration in the solution; it appears to depend strongly on specific structure aspects of the biomolecules and also on the ion extraction conditions in the ion

source. The above observation suggests that these non specific aggregates are predominantly formed post desorption in the expanding plume rather than in solution or upon incorporation into the matrix during solvent evaporation.

Matrix-assisted laser desorption/ionization ion mobility mass spectrometry (MALDI-IM-MS) of DNA

1. Positive ions

All nucleotides listed in Table 8 were tested on a home-built MALDI-IM-TOFMS instrument using positive ion mode for the collision cross section measurements. As we confirmed in previous MALDI-MS experiments that 3-HPA was found to give the best results in the positive oligonucleotide ions, 3-HPA was used as the matrix for all following positive ion mobility experiments of oligonucleotides. The ion mobility experiments for oligonucleotides were conducted at drift cell temperature of 300K, no ion signals corresponding to the protonated oligonucleotides longer than decamer were found. A representative 2D plot of IM-MS conformation space for a mixture of oligonucleotides and peptides is shown in Figure 40. Single, broad peaks are present in all of the arrival time distributions (ATDs) for the protonated oligonucleotides we have examined, suggesting that multiple conformers are present for the protonated oligonucleotides, and the conformers isomerizes as they travel through the drift cell resulting in a "time-averaged" conformation with the broad ATDs.
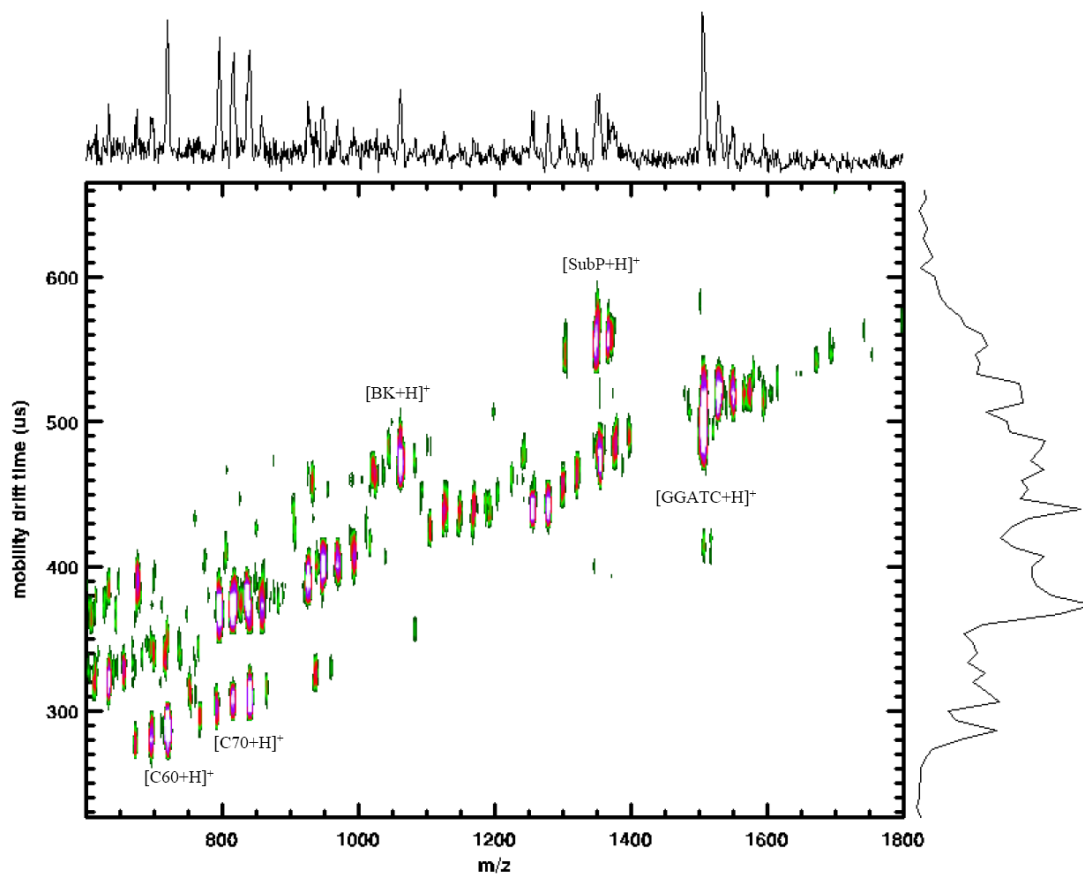
Figure 40. A representative 2D IM-MS plot of conformation space for a mixture of protonated oligonucleotide, peptide ions, and C_{60}.

Table 12. The collision cross sections of the protonated and deprotonated oligonucleotide ions. The error for the measurements is within ±2%.

nt	Squence	[M-H]⁻ m/z	Ω_{exp} (Å²)	[M+H]⁺ m/z	Ω_{exp} (Å²)	[2M+H]⁺ m/z	Ω_{exp} (Å²)
5mer	GGATC	1501.289	244	1503.31	251	3005.60	392
6mer	CTAATC	1749.334	282	1751.35	294		
	GATTAG	1829.35	278	1831.36	298		
	CTGGTC	1781.32	271	1783.34	286		
	GACCAG	1799.35	270	1801.36	294		
8mer	CTCTAGAG	2407.44	322	2409.45	344		
	GATTAGCA	2431.45	329	2433.47	348		
10mer	CACATGGCTG	3025.54		3027.55	406		
	ATCGATCGAT	3024.54	355	3026.56	408		

For each oligonucleotide, a series of ATDs were measured at several different drift voltages. From these measurements, the ion-neutral collision cross sections of the ions are determined using equtation 1 and are calibrated with $[M + H]^+$ ions of bradykinin ($\Omega_{meas} = 245$ Å^2) and substance P ($\Omega_{meas} = 292$ Å^2) [35]. The collision cross sections of the protonated oligonucleotides are shown in Table 12. The cross sections listed in Table 12 are the average of nine measurements over several months and the errors are the relative standard deviations. A summary of these data is illustrated in Figure 41. In Figure 41, it can be seen that the positive oligonucleotide ions have significantly smaller collision cross sections than positive ions of peptides of the same molecular weight. In principle, the difference between the peptide and oligonulceotide cross sections indicates the gas-phase packing efficiency differs for each molecular class. It was also found that the nucleotides of the same length have similar measured collision cross sections, regardless of sequence. For example, all four hexamers have close values in the measured cross sections, even though dCTAATC positively deviates from the average oligonucleotide trendline, but the deviation is less than 3%, suggesting that the hexamers may have similar conformation as well. Generally, the conformational information of the ions can be generated by comparing the experimental cross sections to the calculated cross sections of theoretical structures obtained from molecular modeling. The detail will be presented in ***Theoretical study of DNA***.

Oligonucleotides, peptides, and C60
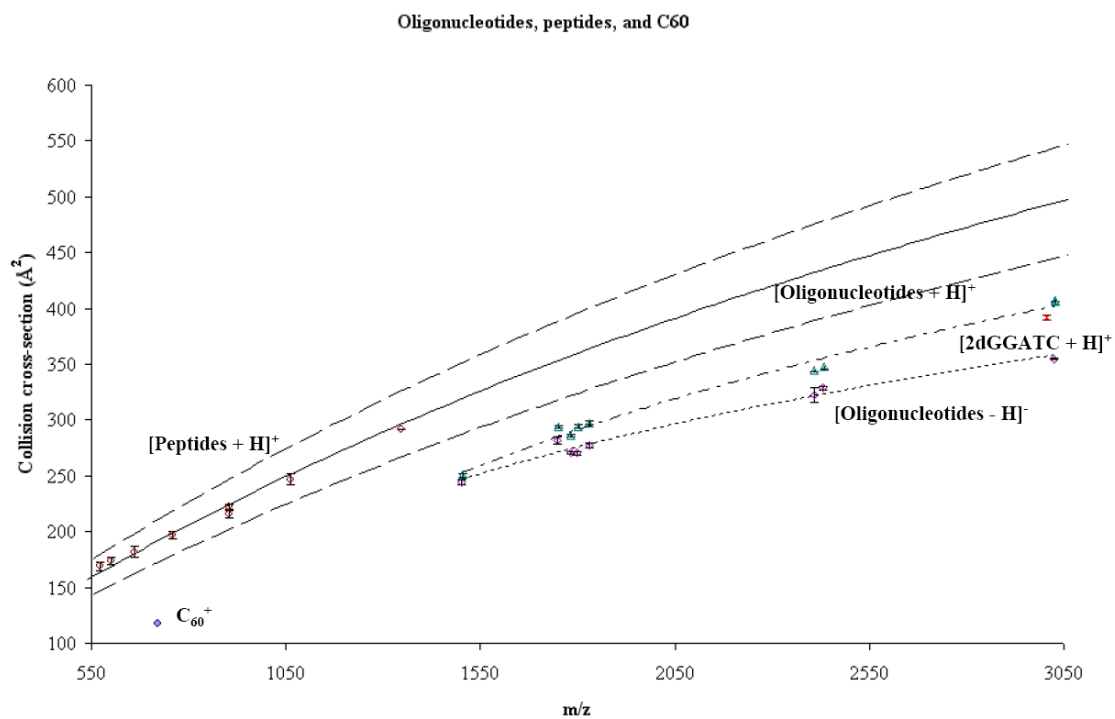
Figure 41. A plot of collision cross section as a function of m/z for different molecular classes including oligonucleotide, peptide ions, and C_{60}.

2. Negative ions

All nucleotides listed in Table 8 were also tested on MALDI-IM-TOFMS instrument using negative ion mode for the collision cross section measurements. As we proved in previous MALDI-MS experiments that 2, 5-DHB with ammonium citrate was the most effective matrix for the detection of negative oligonucleotide ions, it became the standard matrix for all the negative ion mobility experiments of oligonucleotides. The negative ion mobility experiments for oligonucleotides were performed at the same instrumental conditions as the positive mode and gave pretty much the same results as the positive mode as well. No ion signals corresponding to the deprotonated oligonucleotides longer than decamer were found. A representative 2D plot of IM-MS conformation space for a mixture of oligonucleotides and peptides is shown in Figure 42. Single peaks were found for all the deprotonated oligonucleotides, however, the peaks are much narrower than those obtained at positive modes for the same species. It suggests that the deprotonated oligonucleotides may have less multiple conformers than the protonated oligonucleotides. We will discuss it in detail later. The collision cross sections of the deprotonated oligonucleotides are shown in Table 12. The cross sections listed in Table 12 are the average of five measurements over three months and a summary of these data are illustrated in Figure 41 as well. It can be seen that negative ions of oligonucleotides have similar average cross sections trendline to the positive oligonucleotide as the molecular weights increase. However, the negative oligonucleotide ions have slightly smaller collision cross sections than positive ions of oligonucleotides of the same sequence.
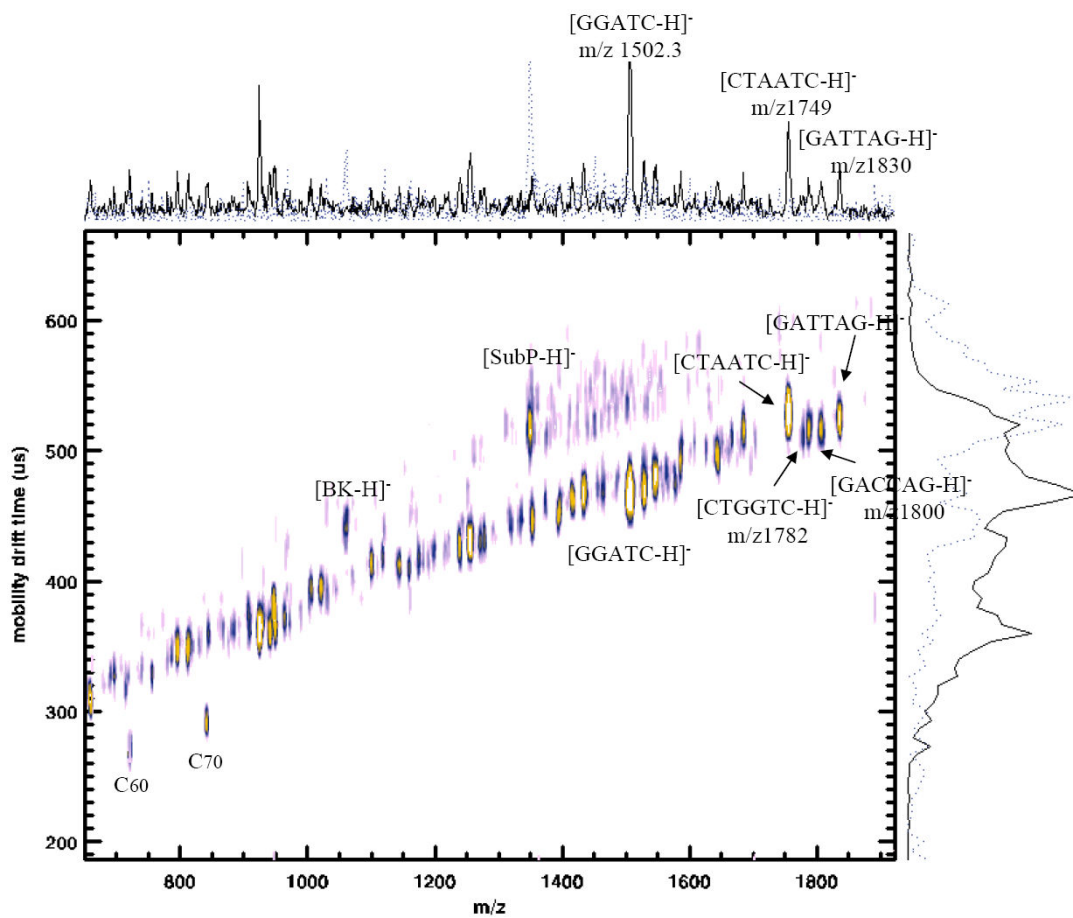
Figure 42. A representative 2D IM-MS plot of conformation space for a mixture of deprotonated oligonucleotide, peptide ions, and C_{60}.

In general, negative ions adopt more than positive ions because there is more charge solvation involved in the folding of negative ions than that of positive ions. Same, the detail conformational information will be presented in the chapter of ***Theoretical study of DNA***.

3. Study of non-covalent complexes for MALDI-IM-MS of DNA

As mentioned previously, noncovalent dimers were observed in most oligonucleotides when they were tested with 3-HPA on STR. We tried to find such noncovalent dimers of oligonucleotides in ion mobility experiments as well. However, only protonated dimers of dGGATC were detected when using 3-HPA as matrix among all the oligonucleotides we have tested. Even though we tried to optimize the experimental condition, i.e., increased concentration of the analyte, adjusted the molar ratio of analyte and matrix, and tuned the instrumental parameters, we were no luck to get better results. Actually, it is easy to understand this observation. Such noncovalent complexes have very weak intermolecular interactions; they are not strong enough to survive under several torr pressure of collision region during the ion mobility experiment. A representative 2D plot of IM-MS conformation space for the protonated dimers of dGGATC is shown in Figure 43. It can be seen the signal for the protonated dimers is not strong, and the peak is broad. The ion-neutral collision cross section of the protonated dimers of dGGATC is determined from ATDs as 392 $Å^2$, which is around 3% smaller than the average trendline of the protonated oligonucleotides (shown in Figure 41). Since the cross section value of the dimers falls within the average trandline
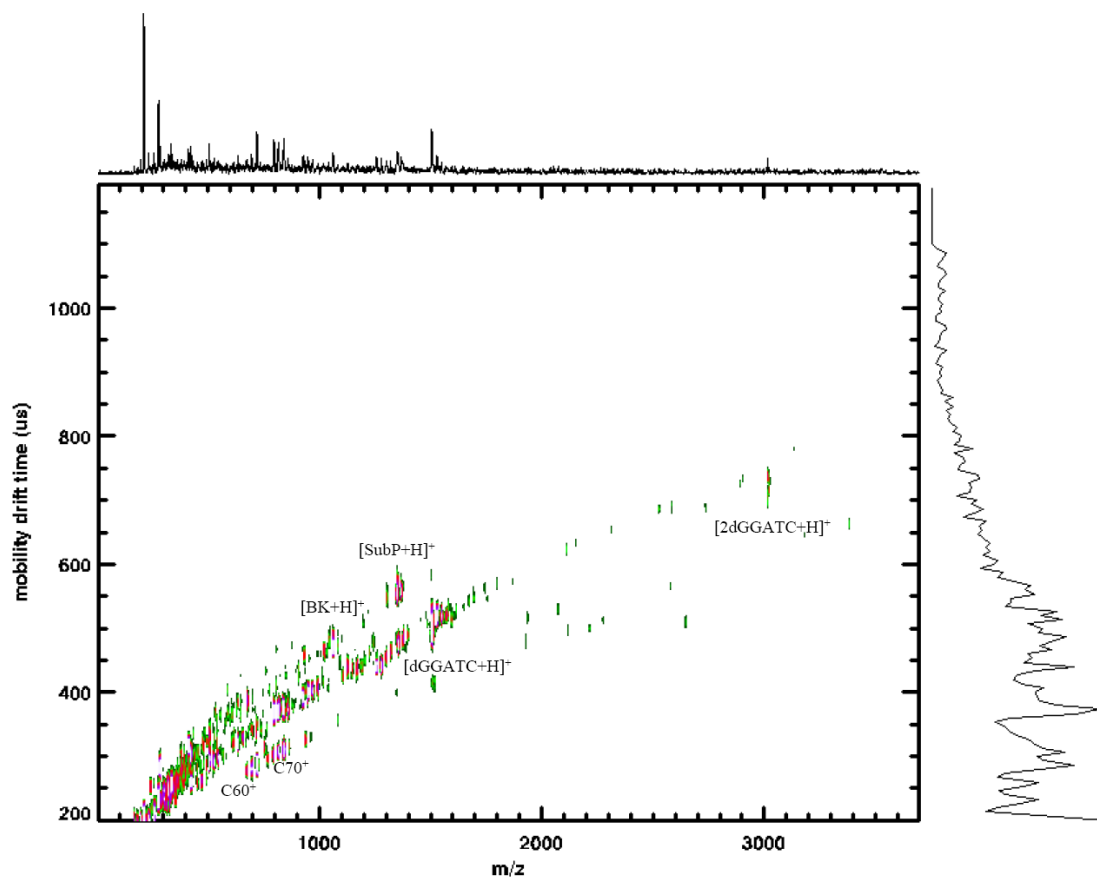
Figure 43. A representative 2D IM-MS plot of conformation space for a mixture of protonated monomer and dimer of dGGATC.

of the protonated oligonucleotides, suggesting that the single strand of dGGATC may keep similar conformation as they form noncovalent complex.

Theoretical study of DNA

Ion mobility spectrometry (IMS) combined with molecular dynamics (MD) has proven to be the most efficient technique for conformational analysis of gas phase structures of biomolecules by measuring the ion-neutral collision cross section of molecular ions [51, 54]. Generally, the conformational information of the ions can be generated by comparing the experimental cross sections to the calculated cross sections of theoretical structures obtained from molecular dynamics. However, the technical challenge in simulation is to generate all the thermodynamically accessible structures that correspond to the real experimental condition. The standard simulations are typically limited to tens of nanoseconds, whereas the molecular ions of ion mobility experiments experience structural changes over millisecond timescale [126, 127]. Two methods are proposed to overcome this limitations: 1) a simulated annealing MD with cluster analysis [135], 2) a generalized non-Boltzman sampling MD with free energy analysis follow by cluster analysis [81]. Method 1 is based on the generation of a pool of initial structures using a simulated annealing process, followed by the minimization of each structure. All the minimized structures are compared using cluster analysis method and different structural families are generated. We used this method to describe the gas phase conformation for peptide ions (detailed in previous chapter). This method increases the probability of generating structures closely related to the most stable

conformers; however, all the resulting structures are minimized at 0 K, which is not the real effective ion temperature. The method is not able to provide the desired thermodynamic information as well. As compared to method 1, method 2 can sample the total conformational space over a wide energy range by using a generalized (non-Boltzman) distribution functions at multiple temperatures [126]. The approach uses a biased potential for the MD simulations generated from the generalized distribution function that, when reweighted correctly, can lead to the desired thermodynamic information. The resulting structures are extracted from free energy minima out of the total conformational space. The method has been proven to be efficient in free energy simulations of peptide ions, thus it was used to study the gas phase structures of oligonucleotide ions.

Another major issue in simulating nucleotide ions is that there is no good force field parameter available for nucleotide fragments. The default nucleotide fragments available in current AMBER9 are the units containing a deprotonated phosphate backbone with a neutralized base. To generate singly charged oligonucleotides, we have to protonate the base or neutralize the phosphate group. A new force field topology database of RESP atomic charge values was developed for modeling singly charged oligonucleotides. RESP charge derivation for new nucleotide fragments was performed using Dimethylpohosphate (conformation gauche, gauche), neutralized Dimethylphate (conformation gauche, gauche), the 4 neutral DNA nucleosides [Deoxyadenosine, Deoxycytidine, Deoxyguanosine and Deoxythymidine] (conformations C2'endo and C3'endo), and the 3 protonated DNA nucleosides [Deoxyadenosine, Deoxycytidine and

Deoxyguanosine] (conformations C2'endo and C3'endo). All geometries were optimized using the HF/6-31G* theory level and four molecular orientations for each optimized geometry were involved in the charge fitting procedure to yield reproducible atom charge values. Two inter-molecular charge constraints between the methyl group of Dimethylpohosphate and the HO3' and HO5' hydroxyls of the target nucleotides were used during the fitting step allowing the definition of the required molecular fragments. Inter-molecular charge equivalencing between the nucleosides were used as well to make the charge values of the deoxyribose atoms (excluding the C1' and H1' atoms) equivalent. The charge derivation procedure was automatically carried out using the R.E.D-IV program[125] and the central, 5'-terminal and 3'-terminal fragments of a nucleotide were simultaneously generated in a single charge derivation. Both topologies A and B, which present the phosphate group resided either at the position 5' or 3' of the target nucleotides, were obtained in this project. The new charge values are compatible with the Cornell *et al.* AMBER force field and proved to be highly reproducible. All data are available at http://q4md-forcefieldtools.org/REDDB/projects/F-83/.

1. dGGATC [M – H]⁻ ion

The singly deprotonated ions of dGGATC was theoretically modeled assuming a zwitterionic structure, where the two guagnine (G) are protonated at N7 and adenine (A) is protonated at N1 while the four phosphate groups are deprotonated (shown in Figure 44).
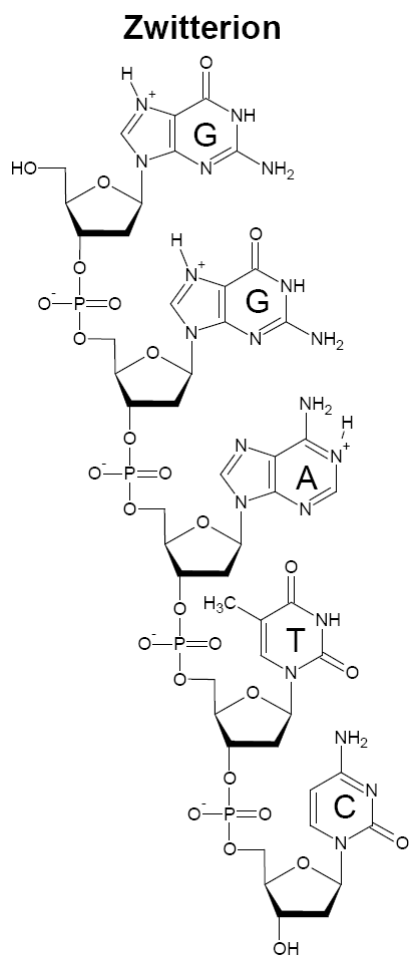
Figure 44. Diagram of the zwitterionic form of dGGATC [M – H]⁻ ions.

Extended structure was used as initial structure for a 500-step minimization and a 20 ps heating step from 0 K to 300K. Then, the enhanced sampling simulation was conducted at 300 K with 300 βvalues exponentially distributed from 250 to 1000 K to generate a much wider energy distribution. After 20 ns simulations, a set of converged weighing factors for different temperatures (i.e., different β) was obtained. Using this set of weighting factors, ten independent enhanced sampling simulations starting from initial structures were conducted for 1 ms, respectively. MD simulations as described above yielded a total of 10,400,000 trajectories for [dGGATC - H]⁻ ions. Figure 45a contains the evolution of the potential energy distribution and the backbone RMSD of all trajectories during the MD simulation time. A correlation between the variation of the potential energy and the backbone RMSD over time was found, suggesting that the conformational inter-conversion barriers are being overcome several times during the MD simulation time. The procedure broadened the energy distribution as well. With the fixed biased potentials, a largely uniform energy distribution was obtained in a wide range of ~500 kcal/mol, compared to the ~100 kcal/mol energy range sampled by the normal MD simulation. The free energy vs. the backbone RMSD distribution was computed based on the probability distribution of all conformations (shown in Figure 45b), and threr free energy minima were observed at the RMSD range of 0.85 to 1.55 (A), the RMSD range of 1.68 to 1.91 (B), and the range of 3.96 to 4.44 (C), respectively. Since the sample size of 10,400,000 trajectories is too huge to be analyzed, 52,000 trajectories were randomly picked for further analysis.
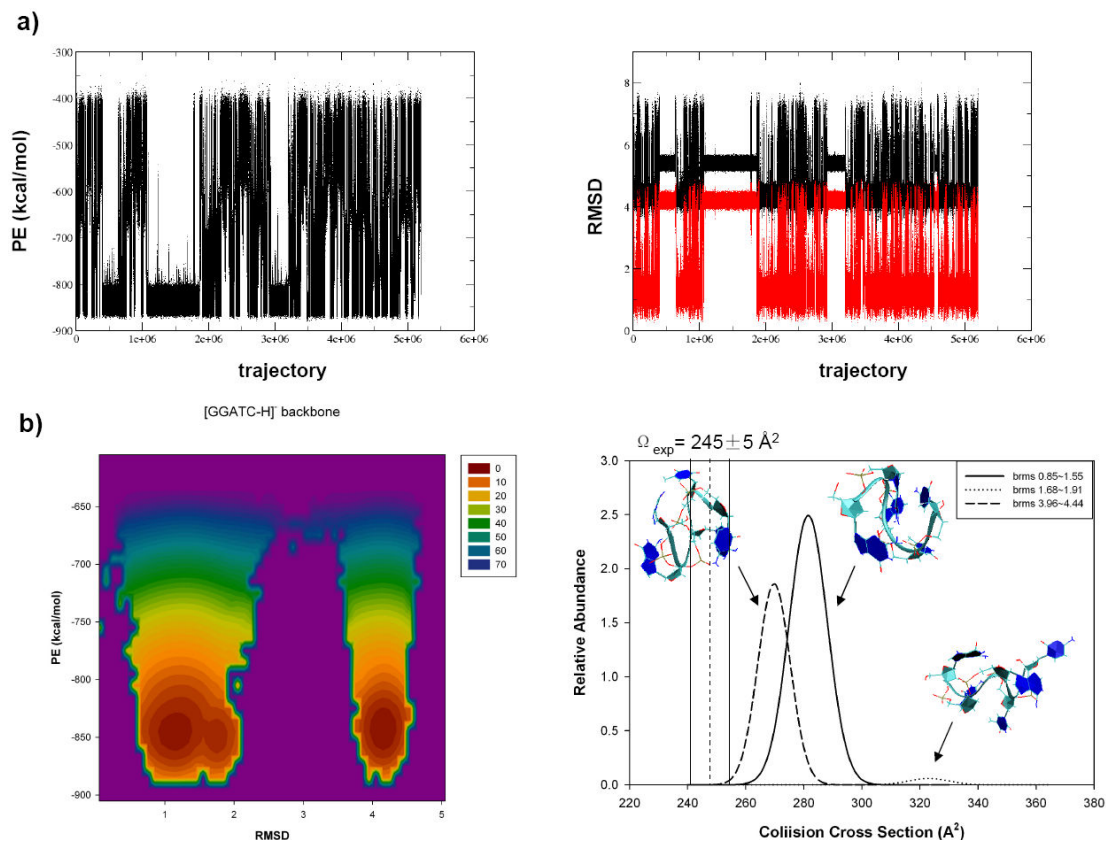
Figure 45. Enhanced sampling MD results determined for dGGATC⁻ ions. a) the evolution of the potential energy distribution and the backbone RMSD of all trajectories during the enhanced sampling MD simulation time; b) the free energy profile as a function of backbone RMSD and potential energy (left), the cross section profile calculated for the lowest free energy structures (right).

There are 25,877 out of 52,000 trajectories fall within A, 2,511 out of 52,000 trajectories fall within B, and 14,659 out of 52,000 trajectories fall within C. Cluster analysis [131] was applied to the trajectories of each region, and only one dominant cluster (population over 75%) was found for each free energy minima region. These structures are characteristic of random coil like a Z-shape for A, a W-shape for B, and a S-shape for C, which are stabilized by the positively charged base units and negatively charged phosphate groups (see in Figure 45b). The collision cross section of the structures in A has been calculated as 281 ± 7 Å^2, the structures in B has the calculated cross section of 323 ± 7 Å^2, and C has the calculated cross section of 269 ± 6 Å^2, which are all larger than the experimental value for [dGGATC – H]$^-$ ions. Statistically, less than 1% structures that fall within C will have the calculated cross section close to the experimental value, but no structures from A and B will match the experimental cross section of [dGGATC – H]$^-$ ions. In the conformer that corresponds to the experimental cross section, no stacked base pairs were observed. The guanine at 5' term hydrogen bonds to both the other G and the thymine, while the adenine is tilt toward the cytosine at 3' term.

2. dGGATC [M + H]$^+$ ion

The situation for simulating singly protonated ions of dGGATC is more complicated than the deprotonated ions of dGGATC. There are three theoretical models available for [dGGATC + H]$^+$, one zwitterionic structure and two non-zwitterionic structures, shown in Figure 46.
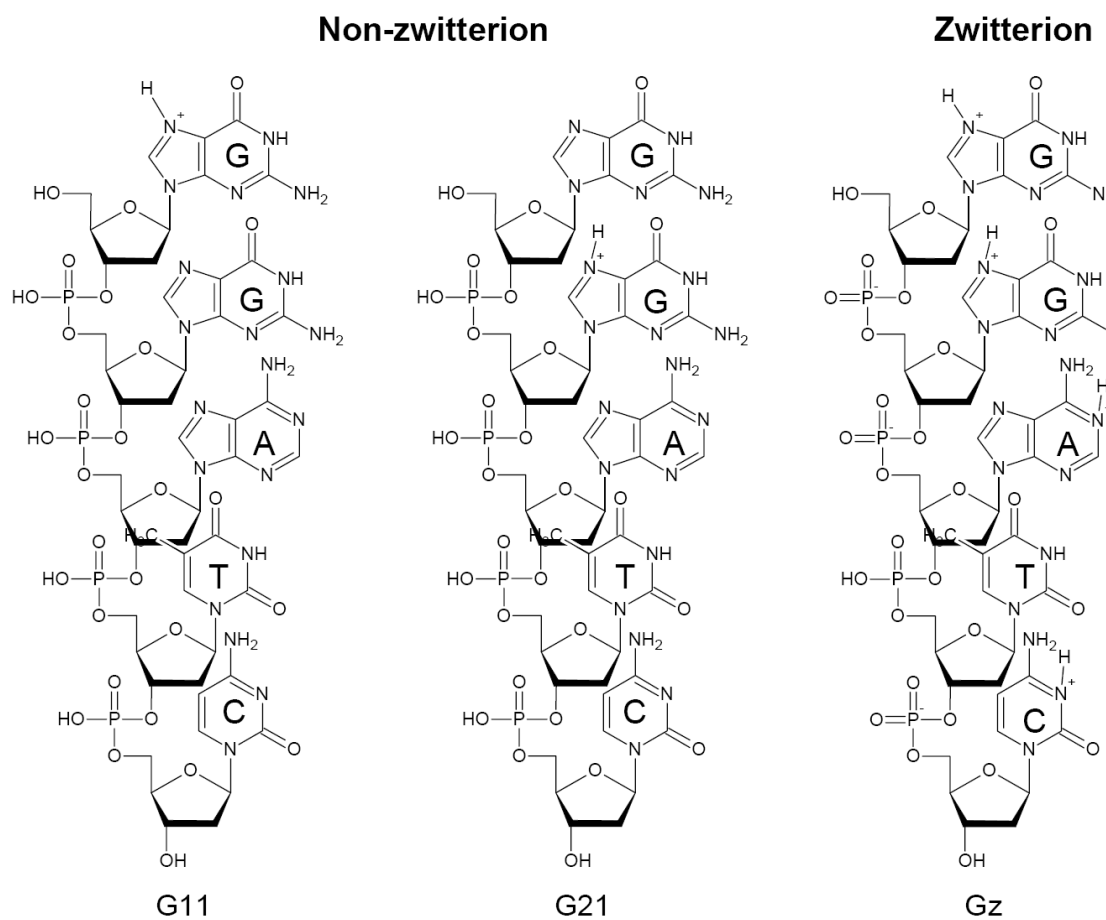
Figure 46. Diagram of the zwitterionic and non-zwitterionic forms of dGGATC$^+$ ions.

In the zwitterionic form, the two guagnine (G) are protonated at N7, adenine (A) is protonated at N1, cytosine (C) is protonated at N3, and only three of the phosphate groups are deprotonated. In the non-zwitterionic forms, all the phosphate groups are neutralized and one of either G is protonated at N7. To differentiate the models, the zwitterionic form is called Gz, the non-zwitterionic form with the protonated G at 5' term is called G11 and the other is called G21.

For each case, extended structure was used as initial structure, a total of ~4,800,000 trajectories were generated after the enhanced sampling simulation. The evolution of the potential energy distribution and the backbone RMSD of all trajectories during the MD simulation time for each case are presented in Figure 47a – 49a and the free energy vs. the backbone RMSD distribution are shown in Figures 47b – 49b. For Gz, only one free energy minima was observed at the backbone RMSD range of 2.5 to 3 which is dominated by one structure element based on the cluster analysis result. The structure of Gz is characterized by a Z-shape, similar to the structure observed for dGGATC$^-$ ions and havs a calculated cross-section value of 273 ± 6 Å^2. For G11, one free energy minima was observed at the backbone RMSD range of 2.3 to 3.3 and one dominated structure was found within that region. The structure is an elongated form with the calculated cross section of 308 ± 5 Å^2. For G21, three free energy minima were observed at the backbone RMSD of ~4, RMSD of ~4.5, and RMSD of ~5 and each case was dominated by one structure.
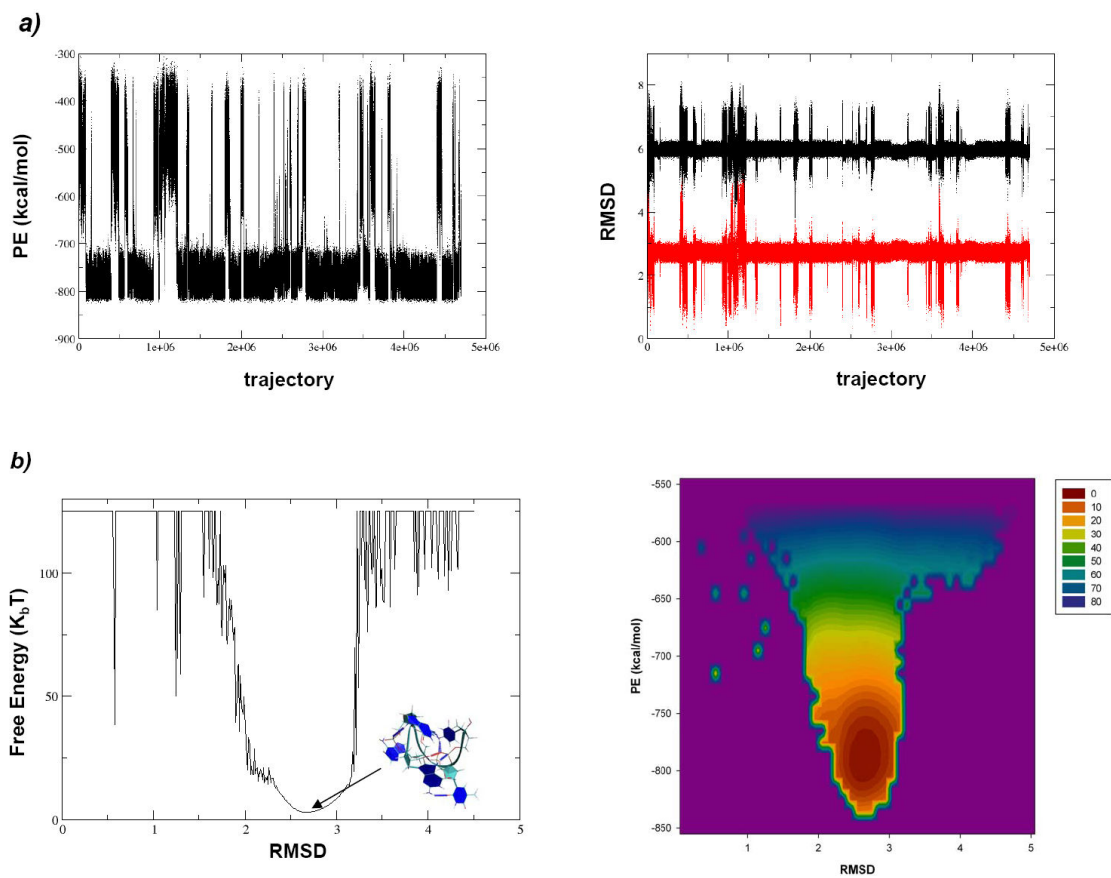
Figure 47. Enhanced sampling MD results determined for the zwitterionic form of dGGATC$^+$ ions. a) the evolution of the potential energy distribution and the backbone RMSD of all trajectories during the enhanced sampling MD simulation time; b) a 2D free energy profile as a function of backbone RMSD (left), a 3D free energy profile as a function of backbone RMSD and potential energy (right).
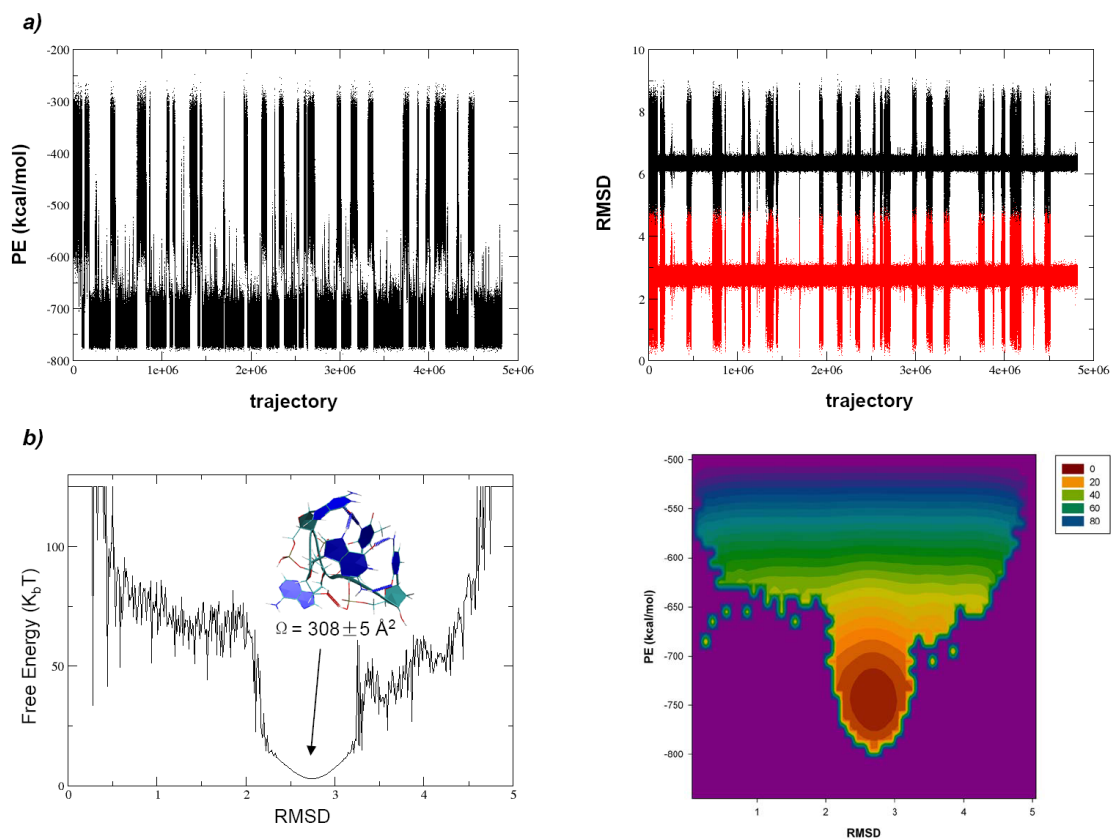
Figure 48. Enhanced sampling MD results determined for the nonzwitterionic form of dGGATC$^+$ ions with the charge-giving proton at G1. a) the evolution of the potential energy distribution and the backbone RMSD of all trajectories during the enhanced sampling MD simulation time; b) a 2D free energy profile as a function of backbone RMSD (left), a 3D free energy profile as a function of backbone RMSD and potential energy (right).
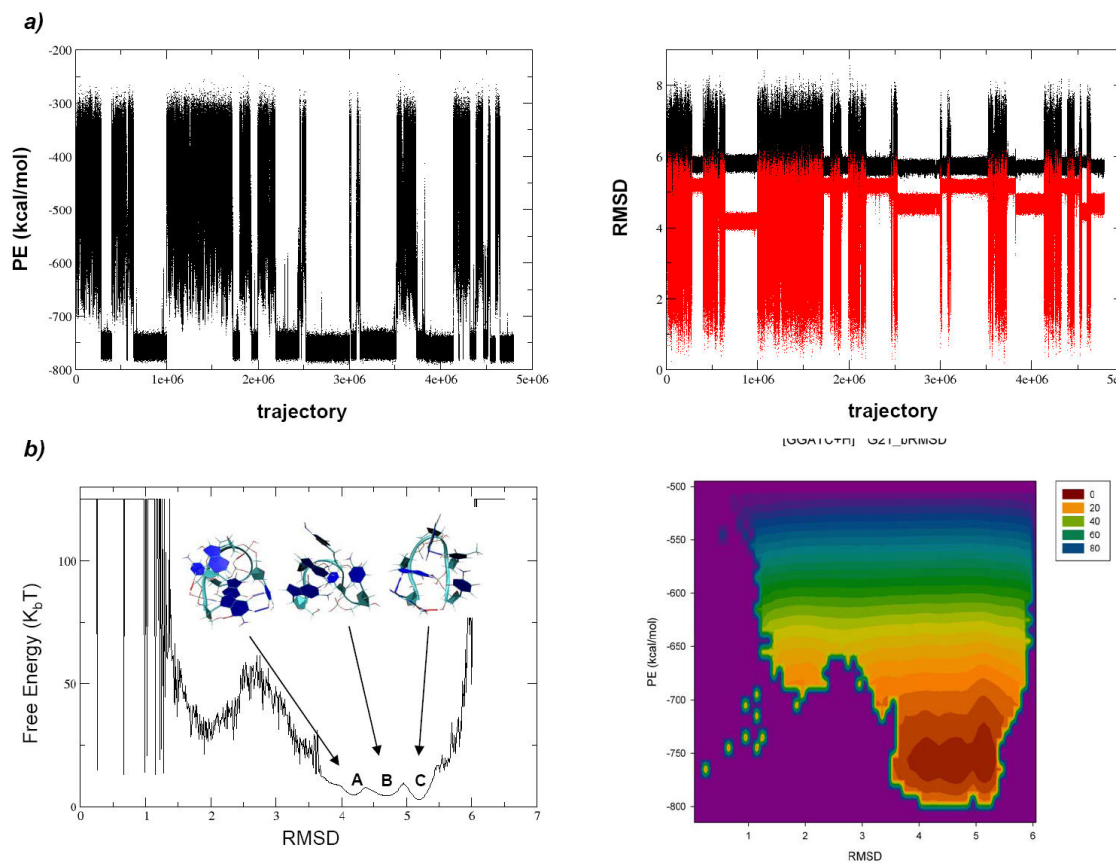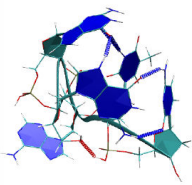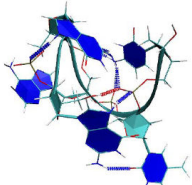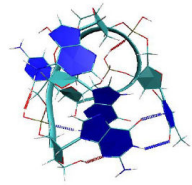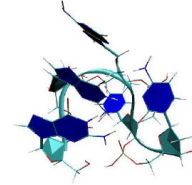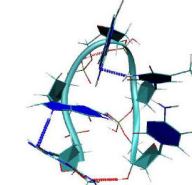
Figure 49. Enhanced sampling MD results determined for the nonzwitterionic form of dGGATC$^+$ ions with the charge-giving proton at G2. a) the evolution of the potential energy distribution and the backbone RMSD of all trajectories during the enhanced sampling MD simulation time; b) a 2D free energy profile as a function of backbone RMSD (left), a 3D free energy profile as a function of backbone RMSD and potential energy (right).

The structure at RMSD 4 is a ρ-shape form with the calculated cross section of 268 ± 5 Å^2; while the structure at RMSD 4.5 is characterized by a hairpin with the calculated cross section of 256 ± 5 Å^2; while the structure at RMSD 5 is a U-shape form with the calculated cross section of 267 ± 5 Å^2. The representative structures classified from each free energy minima and the probability of each structure that matches the experimental cross section value for [dGGATC + H]$^+$ ions are shown in Figure 50. A total of 50% of the Ω-shape structure with the protonated G at second position are expected to have the cross section values that match the experimental cross section of [dGGATC + H]$^+$ ions. The rest of structures all contribute slightly to the total conformational space of [dGGATC + H]$^+$ ions except the structures from G11. The ATDs observed for [dGGATC + H]$^+$ ions can be the "average" of those lower free energy conformers interconverting during the IM experimental timescale, which might be the reason for the broader ATDs of dGGATC [M + H]$^+$ ions than those of dGGATC$^-$ ions.

3. *dGGATC [2M + H]$^+$ ion*

To avoid intermolecular charge repulsion between two strands of dGGATC, all the phosphate group at backbone are neutralized, thus only non-zwitterionic structures was assuming as the theoretical model for protonated dimers of dGGATC , one with the protonated G at 5' term on one strand (G11-G0), and the other with the protonated G at 2nd position (G21-G0).

a)

	A	B	C	D	E
	G1	GZ	G2	G2	G2
Ω	308.2 Å²	272.9 Å²	268.5 Å²	255.7 Å²	267.4 Å²
σ	5.4 Å²	5.5 Å²	4.7 Å²	4.6 Å²	4.9 Å²
P	0.00%	0.11%	0.37%	49.5%	1.10%

b)
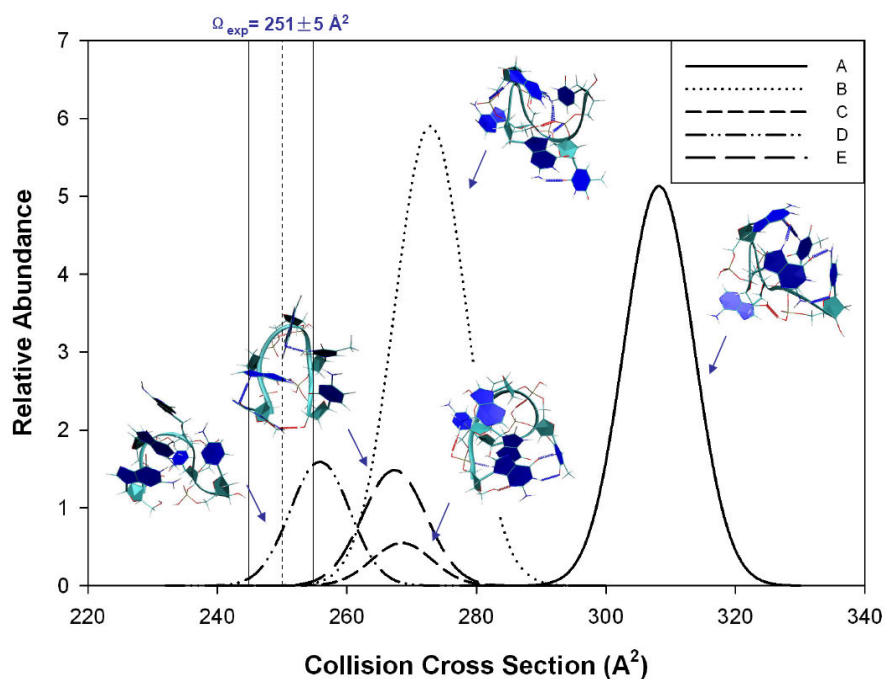
Figure 50. a) The representative structures classified for each free energy minima and the probability of each structure that matches the experimental cross section of dGGATC$^+$ ions; b) the cross section profile for the lowest free energy structures.

Since no dimers of dGGATC can be survived when the simulated temperature passes over 600K, we are unable to use enhanced sampling method to simulate the gas phase structures of the protonated dimmer of dGGATC ions. For each case, β-helix was used as initial structure for a 500-step minimization and a 20 ps heating step from 0 K to 300K. Then, the normal MD simulation was conducted at 300 K for a timescale of 10 ns generating 20,000 trajectories. The cross section of each trajectory was calculated and a scatter plot of cross section versus energy was generated to help compare theory with experiment. In both cases, the theoretical cross section values are way larger than the experimental ones. Unfortunately, no reasonable conformer was found for matching the protonated dimers of dGGATC under experimental condition. The normal MD simulation only sampled at a narrow energy range of ~100 kcal/mol, it was unable to generate all the thermodynamically accessible structures that correspond to a certain experimental condition, which can explain for the failure of obtaining the desired results.

Conclusion

A series of different length oligonucleotids were tested by MALDI-MS on both positive and negative ion mode. 3-HPA in combination with ammonium citrate has shown to be the best matrix for detection of positive oligonucleotide ions, while 2,5-DHB with ammonium citrate is the preferred matrix to measure deprotonated oliognucleotides. Noncovalent dimmer ions of oligonucleotides were also observed for most oligonucleotide ions when mixtures of complementary or noncomplementary strands were analyzed with 3-HPA via positive MALDI-MS. However, non-specific

duplex was observed for the short-length oligonucleotides we have examined. IMS as the most efficient separation method predicted an average cross section trendline for both positive and negative oligonucleotide ions under gas phase. The singly protonated oligonucleotide ions have significant smaller collision cross sections than positive ions of peptides of the same molecular weight, but have slightly larger cross section values than negative ions of oligonucleotides of the same species. The measured collision cross sections of all the oligonucleotide ions we have tested in both positive and negative ion mode fall on the average cross section trendline of oiligonucleotides, suggesting that they may have similar conformation. Further understanding of the structures within conformation space might be achieved through molecular dynamic simulations. A novel generalized non-Boltzman sampling MD has proven to be efficient in free energy analysis was used for theoretical study of gas phase conformation of oligonucleotide ions. In the case of dGGATC$^-$ ions, a Z-shape zwitterion form might be the possible conformer existed in the gas phase. On the other hand, several lower free energy structures were found to contribute to the conformational space of dGGATC$^+$ ions. It might be the reason for the broader ATDs of dGGATC$^+$ ions than those of dGGATC$^-$ ions. Overall, the method provides a higher confidence in the gas phase structure identification of oligonucleotides when compared to the experimental determined cross section values. The study described here is just the first validation of combining IMS data with theoretical tools for understanding the gas phase conformation of oligonucleotides under certain experimental conditions. Based on this work, more

complicated and systematic oligonucleotides can be explored to further understanding the folding mechanism of biomolecular ions in the gas phase.

CHAPTER V

CONCLUSION

In this study, we have shown that the combination of ion mobility with mass spectrometry affords a powerful approach to explore the gas phase structures of biomolecule ions by measuring collision cross sections of mass identified ions. The technique is particularly useful in conjunction with molecular modeling (e.g. molecular mechanics or molecular dynamics) and bioinformatics tools (i.e. cluster analysis).

Gas-phase structures of polypeptide ions

A database of ion-neutral collision cross-sections for hundreds of $[M + H]^+$ proteolytic peptide ions has been built using MALDI-IM-TOFMS, containing structure and mass information. The majority (63%) of the peptide ion collision cross-sections correlate well with structures that are best described as charge-solvated globules, but a significant number of the peptide ions exhibit collision cross-sections that are significantly larger or smaller than the average, globular mobility-mass correlation. Molecular simulations results suggest that peptide ions having larger than predicted collision cross-sections have more open conformations and can be good candidates for helical structures. Whereas the peptides with smaller cross-sections would prefer more compact structure due to the salvation interaction from their polar residues. When compared to the collision cross-sections of singly charged peptide ions measured by ESI, we find some difference between the measured values. Even though it is not known in detail whether there are systematic differences between the different ionization methods,

simulation results indicate that they might yield gas-phase ions which can differ in terms of structure or conformation. For short peptides with up to ~ 7 residues, the proposed ion structures for ESI and MALDI have similar backbone structures and the different projection of side chains cause the difference in collision cross sections. With the increase sequence, peptide ions have more degrees of freedom for positioning structural elements in backbone; both side chain orientation and backbone structure can be the reasons for the difference in collision cross sections between ESI and MALDI.

At the same time, a novel cluster analysis method has been developed for extracting structure information from simulated annealing results with higher statistical confidence. It evaluates candidate structures from an ensemble of many conformations rather than the lowest energy structure, which better represents the peptide ion structure under real experimental conditions. In the case of MIFAGIK peptide ion, cluster analysis suggests that $[M + H]^+$ ions formed by MALDI or ESI both prefer a α-turn structure; this structural preference is probably the result of stabilization afforded by charge-solvation by the backbone amide groups. That is, there appears to be a significance preference for helical or partial helical conformers by the low energy ions, but the major fraction of the ion populations exists as α-turn, random coil, and β-turn conformers. Although the derivatives of MIFAGIK which influence charge site and intra-molecular interactions alter the distribution of various conformers, the most dramatic changes are observed for the $[M + Na]^+$ ion, which show a strong preference for random coil conformers owing to the strong solvation by the backbone amide groups. We tested the 'sensitivity' of the cluster analysis method by introducing subtle variations in peptide ion composition as

well. For example, the N-acetyl and methylester derivatives as well as the Boc and Fmoc derivations all share very similar backbone structures. All the results presented here support the hypothesis that difference in collision cross-sections for MIFAGIK peptide ions are related to the different orientation of side chains.

Gas-phase structures of oligodeoxynucleotide ions

A series of different length oligonucleotids were tested by MALDI-IM-TOFMS in both positive and negative ion mode, which predicted an average cross section trendline for both positive and negative oligonucleotide ions in the gas phase. A novel generalized non-Boltzman sampling MD has been utilized to generate a conformational space for the give molecule ions with the desired thermodynamic information. The representative structures can be extracted out of the total conformational space by the free energy values. In the case of dGGATC ion, inspection of the representative structure showed that the singly charged negative ions mostly adopt a Z-shape zwitterion form while the singly charged positive ions can exist in several different backbone structures in both canonical and zwitterion form in the gas phase. These results are in good agreement with the experimental observations.

Future directions

Overall, the aim of our research at further development of methodology for extracting structural information for gas-phase biomolecule ions, which can contribute to better understanding of the conformational space occupied by biomolecules are of

considerable interest. The work presented in the paper is just the first step in this direction.

Future studies will extend to IM-MS structural characterization to larger, more complex biomolecules, i.e., melittin, the principal active protein found in bee venom. Melittin (GIGAVLK^{7}VLTTGLPALISW-IK^{21}RKR^{24}QQ-NH$_2$) [136] is of interest because it exists as unstructured monomer in aqueous phase; however, in basic or high ionic strength solutions melittin aggregates to form tetramers with helical characteristics present in each component [137, 138]. Our previous MD simulations suggest that as the solvent dielectric constant increases (from methanol to ethanol, propanol, butanol, and finally to water), helical melittin monomers (all basic sites are protonated, +5 or +6) undergo conformational changes to highly disordered structures and that the C-terminus is more stable than the N-terminus. Previous mass spectrometry fragmentation studies were interpreted as evidence that gas-phase melittin ions are helical [139].

Cluster analysis of the various charge states of a peptide as large as melittin may increase our confidence of the tentative assignment of the backbone structure. A possible improvement to the method which could easily be extended to larger proteins would involve digestion of the protein into peptide subunits to be interrogated individually by applying the cluster algorithm. The peptide subunits are more tractable for both MD simulations and cluster analysis, and may retain structural preferences of the whole protein. As a specific example, trypsin digests of melittin would yield GIGAVLK^7 and VLTTGLPALISWIK^{21} as well as fragments containing R^{22}, K^{23}, R^{24} and these fragment ions could be examined separately, including Ac- and methyl ester

forms, from the intact peptide. Statistical methods will need to be developed to aid in assembling results from the peptide subunits.

The backbone of a peptide can be represented as the (x, y, z) coordinates of an amino acid's four heavy atoms (N, C$_\alpha$, C, and O) along the backbone. The root-mean-square-deviation (RMSD) between two optimally-aligned peptides gives a measure of dissimilarity among peptides. These pairwise distances were used for the distance-based clustering procedure in our current clustering method. Conversely, a peptide's backbone can be described by dihedral angle pair (φ, ψ) along the backbone as well. The similarity of the vectors of dihedral angle pairs among peptides can also be used in a model-based clustering procedure. The clustering method can be extended to use the dihedral angles along a backbone as well.

The study of specific non covalent complex of oligonucleotides will be particularly interesting. The H-bonding interactions in nucleic acids play a crucial role in the double helical structure of DNA. Theoretical calculation on the G-C and A-T pairs provides the required information about the strength of H-bonding in these systems and the respective binding energies are 20 and 17 kcal/mol. DNA with high GC-content should be more stable than DNA with high AT-content owing to the contribution of π-stacking interactions. For example, the mixtures of GATTAG with CTAATC and CTGGTC with GACCAG, the bases on one strand are all complementary to the bases on the other strand. As the percentage of GC base pairs determines the strength between the two strands of DNA, the duplex formed by CTGGTC and GACCAG are stronger-interacting strands than the strands between GATTAG and CTAATC. The strands with

stronger interaction may more easily form high order structures such as double helix, but the weaker interacting strands can be easily pulled apart and form random structures. However, the duplex may not survive under MALDI; ESI will be the better choice to study ion structure of DNA non-covalent complex. Now the mobility instrument with both ionization sources is under development in our lab.

REFERENCES

1. Drenth, J. *Principles of Protein X-ray Crystallography*; Springer: New York, 1994.

2. Bax, A. Multidimensional nuclear magnetic resonance methods for protein studies *Current Opinion in Structural Biology* **1994**, *4*, 738-744.

3. Fenn, J.; Mann, M.; Meng, C.; Wong, S.; Whitehouse, C. Electrospray ionization for mass spectrometry of large biomolecules *Science* **1989**, *246*, 64-71.

4. Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers *Analytical Chemistry* **2008**, *63*, 1193A-1203A.

5. Przybylski, M.; Glocker, M. O. Electrospray mass spectrometry of biomacromolecular complexes with noncovalent interactions - New analytical perspectives for supramolecular chemistry and molecular recognition processes *Angewandte Chemie International Edition in English* **1996**, *35*, 806-826.

6. Karas, M.; Bahr, U. Laser desorption ionization mass spectrometry of large biomolecules *TrAC Trends in Analytical Chemistry*, *9*, 321-325.

7. Biemann, K.; Scoble, H. Characterization by tandem mass spectrometry of structural modifications in proteins *Science* **1987**, *237*, 992-998.

8. Burlingame, A. L.; Carr, S. A., Eds. *Mass Spectrometry in the Biological Sciences*; Humana Press: Totowa NJ, 1996.

9. Ruotolo, B. T.; Benesch, J. L. P.; Sandercock, A. M.; Hyung, S.-J.; Robinson, C. V. Ion mobility-mass spectrometry analysis of large protein complexes *Nature Protocols* **2008**, *3*, 1139-1152.

10. Ruotolo, B. T.; Hyung, S.-j.; Robinson, P. M.; Giles, K. B.; H., R.; Robinson, C. V. Ion mobility-mass spectrometry reveals long-lived, unfolded intermediates in the dissociation of protein complexes *Angewandte Chemie, International Edition* **2007**, *46*, 8001-8004.

11. Ruotolo, B. T.; Giles, K.; Campuzano, I.; Sandercock, A. M.; Bateman, R. H.; Robinson, C. V. Evidence for macromolecular protein rings in the absence of bulk water *Science* **2005**, *310*, 1658-1661.

12. Uetrecht, C.; Versluis, C.; Watts, N. R.; Wingfield, P. T.; Steven, A. C.; Heck, A. J. R. Stability and shape of hepatitis B virus capsids in vacuo *Angewandte Chemie, International Edition* **2008**, *47*, 6247-6251.

13. Loo, J. A.; Kaddis, C. S. Direct characterization of protein complexes by electrospray ionization mass spectrometry and ion mobility analysis *Mass Spectrometry of Protein Interactions* **2007**, 1-23.

14. Kaddis, C. S.; Lomeli, S. H.; Yin, S.; Berhane, B.; Apostol, M. I.; Kickhoefer, V. A.; Rome, L. H.; Loo, J. A. Sizing large proteins and protein complexes by electrospray ionization mass spectrometry and ion mobility *Journal of the American Society for Mass Spectrometry* **2007**, *18*, 1206-1216.

15. Kaddis, C. S.; Loo, J. A. Native protein MS and ion mobility: large flying proteins with ESI *Analytical Chemistry* **2007**, *79*, 1778-1784.

16. Barrera, N. P.; Di Bartolo, N.; Booth, P. J.; Robinson, C. V. Micelles Protect membrane complexes from solution to vacuum *Science* **2008**, *321*, 243-246.

17. Miranker, A.; Robinson, C.; Radford, S.; Aplin, R.; Dobson, C. Detection of transient protein folding populations by mass spectrometry *Science* **1993**, *262*, 896-900.

18. Przybylski, M. Mass spectromtric approaches to the characterization of tertiary and supramolecular structures of biomarcromolecules *Adv Mass Spectrom* **1995**, *13*, 257-283.

19. Wood, T. D.; Chorush, R. A.; Wampler Fm, III; Little, D. P.; O'Connor, P. B.; McLafferty, F. W. Gas-phase folding and unfolding of cytochrome *c* cations *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 2451-2454.

20. Paulin, L.; Corey, R. B.; Branson, H. R. The structure of proteins *Proc. Natl. Acad. Sci. USA* **1951**, *37*, 205-211.

21. Kauzmann, W.; C.B. Anfinsen, J. M. L. A. K. B.; John, T. E. Some factors in the interpretation of protein denaturation *Advances in Protein Chemistry* **1959**, *14*, 1-63.

22. Richards, F. M. The interpretation of protein structures: Total volume, group volume distributions and packing density *Journal of Molecular Biology* **1974**, *82*, 1-14.

23. Jeffery, G. A.; Saenger, W. *Hydrogen Bonding in Biological Structure*; Springer-Verlag: Berlin, 1991.

24. Watson, J. D.; Crick, F. H. C. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid *Nature* **1953**, *171*, 737-738.

25. Suckau, D.; Shi, Y.; Beu, S. C.; Senko, M. W.; Quinn, J. P.; Wampler, F. M.; McLafferty, F. W. Coexisting stable conformations of gaseous protien ions *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 790-793.

26. McLafferty, F. W.; Guan, Z.; Haupts, U.; Wood, T. D.; Kelleher, N. L. Gaseous conformational structures of cytochrome *c Journal of the American Chemical Society* **1998**, *120*, 4732-4740.

27. Campbell, S.; Rodgers, M. T.; Marzluff, E. M.; Beauchamp, J. L. Deuterium exchange reactions as a probe of biomolecule structure. Fundamental studies of cas phase H/D exchange reactions of protonated glycine oligomers with D2O, CD3OD, CD3CO2D, and ND3 *Journal of the American Chemical Society* **1995**, *117*, 12840-12854.

28. Gard, E.; Green, M. K.; Bregar, J.; Lebrilla, C. B. Gas-phase hydrogen-deuterium exchange as a molecular probe for the interaction of methanol and protonated peptides *Journal of the American Society for Mass Spectrometry* **1994**, *5*, 623-631.

29. Green, M. K.; Lebrilla, C. B. The role of proton-bridged intermediates in promoting hydrogen-deuterium exchange in gas-phase protonated diamines, peptides and proteins *International Journal of Mass Spectrometry* **1998**, *175*, 15-26.

30. Eiceman, G. A. Advances in ion mobility spectrometry: 1980-1990. *Critical Reviews in Analytical Chemistry* **1991**, *22*, 471-490.

31. Eiceman, G. A.; Karpas, Z*., Ion Mobility Spectrometry*, 2nd ed.; CRC Press: Boca Raton FL, 2005.

32. St. Louis, R. H.; Hill, H. H. Ion mobility spectrometry in analytical chemistry *Critical Reviews in Analytical Chemistry* **1990**, *21*, 321-355.

33. Karpas, Z.; Stimac, R. M.; Rappoport, Z. Differentiating between large isomers and derivation of structural information by ion mobility spectrometry mass-spectrometry techniques *International Journal of Mass Spectrometry and Ion Processes* **1988**, *83*, 163-175.

34. McLean, J. A.; Ruotolo, B. T.; Gillig, K. J.; Russell, D. H. Ion mobility–mass spectrometry: a new paradigm for proteomics *International Journal of Mass Spectrometry* **2005**, *240*, 301-315.

35.   Stone, E.; Gillig, K. J.; Ruotolo, B.; Fuhrer, K.; Gonin, M.; Schultz, A.; Russell, D. H. Surface-induced dissociation on a MALDI-ion mobility-orthogonal time-of-flight mass spectrometer: sequencing peptides from an "in-solution" protein digest *Analytical Chemistry* **2001**, *73*, 2233-2238.

36.   Stone, E. G.; Gillig, K. J.; Ruotolo, B. T.; Russell, D. H. Optimization of a matrix-assisted laser desorption ionization ion mobility surface induced dissociation/orthogonal time-of-flight mass spectrometer: simultaneous acquisition of multiple correlated MS1 and MS2 spectra *International Journal of Mass Spectrometry* **2001**, *212*, 519-533.

37.   Hoaglund-Hyzer, C. S.; Li, J.; Clemmer, D. E. Mobility labeling for parallel CID of ion mixtures *Analytical Chemistry* **2000**, *72*, 2737-2740.

38.   Hoaglund-Hyzer, C. S.; Lee, Y. J.; Counterman, A. E.; Clemmer, D. E. Coupling ion mobility separations, collisional activation techniques, and multiple stages of MS for analysis of complex peptide mixtures *Analytical Chemistry* **2002**, *74*, 992-1006.

39.   Lee, Y. J.; Hoaglund-Hyzer, C. S.; Taraszka, J. A.; Zientara, G. A.; Counterman, A. E.; Clemmer, D. E. Collision-induced dissociation of mobility-separated ions using an orifice-skimmer cone at the back of a drift tube *Analytical Chemistry* **2001**, *73*, 3549-3555.

40.   Clowers, B. H.; Hill, H. H. Influence of cation adduction on the separation characteristics of flavonoid diglycoside isomers using dual gate-ion mobility-quadrupole ion trap mass spectrometry *Journal of Mass Spectrometry* **2006**, *41*, 339-351.

41.   Badman, E. R.; Myung, S.; Clemmer, D. E. Gas-phase separations of protein and peptide ion fragments generated by collision-induced dissociation in an ion trap *Analytical Chemistry* **2002**, *74*, 4889-4894.

42.   Hilderbrand, A. E.; Myung, S.; Srebalus Barnes, C. A.; Clemmer, D. E. Development of LC-IMS-CID-TOFMS techniques: analysis of a 256 component tetrapeptide combinatorial library *Journal of the American Society for Mass Spectrometry* **2003**, *14*, 1424-1436.

43.   Koomen, J. M.; Ruotolo, B. T.; Gillig, K. J.; McLean, J. A.; Russell, D. H.; Kang, M.; Dunbar, K. R.; Fuhrer, K.; Gonin, M.; Schultz, J. A. Oligonucleotide analysis with MALDI-ion-mobility-TOFMS *Anal. Bioanal. Chem.* **2002**, *373*, 612-617.

44.   Woods, A. S.; Ugarov, M. V.; Egan, T.; Koomen, J. M.; Gillig, K. J.; Fuhrer, K.; Gonin, M.; Schultz, J. A. Lipid/peptide/nucleotide separation with MALDI-ion mobility-TOF MS *Analytical Chemistry* **2004**, *76*, 2187-2195.

45.   Vonhelden, G.; Hsu, M. T.; Kemper, P. R.; Bowers, M. T. Structures of cluster ions from 3 to 60 atoms-linears to rings to fullerenes *Journal of Chemical Physics* **1991**, *95*, 3835-3837.

46.   Vonhelden, G.; Kemper, P. R.; Gotts, N. G.; Bowers, M. T. Isomers of small carbon cluster anions-linears-linear-chains with up to 20 atoms *Science* **1993**, *259*, 1300-1302.

47.   Vonhelden, G.; Gotts, N. G.; Bowers, M. T. Annealing of carbon cluster cations-rings to rings and rings to fullerenes *Journal of the American Chemical Society* **1993**, *115*, 4363-4364.

48.   Vonhelden, G.; Hsu, M. T.; Gotts, N. G.; Kemper, P. R.; Bowers, M. T. Do small fullerenes exist only on the computer-experimental results on C-20(+/-) and C-24(+/-) *Chemical Physics Letters* **1993**, *204*, 15-22.

49.   Jarrold, M. F.; Constant, V. A. Silicon cluster ions-evidence for a structural transition *Physical Review Letters* **1991**, *67*, 2994-2997.

50.   Jarrold, M. F.; Bower, J. E. Mobilities of silicon cluster ions-the reactivity of silicon sausages and spheres *Journal of Chemical Physics* **1992**, *96*, 9180-9190.

51.   Valentine, S. J.; Counterman, A. E.; Clemmer, D. E. A database of 660 peptide ion cross sections: use of intrinsic size parameters for bona fide predictions of cross sections *Journal of the American Society of Mass Spectometry* **1999**, *10*, 1188-1211.

52.   Hudgins, R. R.; Ratner, M. A.; Jarrold, M. F. Design of helices that are stable in vacuo *Journal of the American Chemical Society* **1998**, *120*, 12974-12975.

53.   Hudgins, R. R.; Mao, Y.; Ratner, M. A.; Jarrold, M. F. Conformations of Gly_nH^+ and Ala_nH^+ peptides in the gas phase *Biophysical Journal* **1999**, *76*, 1591-1597.

54.   Tao, L.; McLean, J. R.; McLean, J. A.; Russell, D. H. A collision cross-section database of singly-charged peptide ions *Journal of the American Society for Mass Spectrometry* **2007**, *18*, 1232-1238.

55.   Ruotolo, B. T.; Russell, D. H. Gas-phase conformations of proteolytically derived protein fragments: influence of solvent on peptide conformation *Journal of Physical Chemistry B* **2004**, *108*, 15321-15331.

56.   Ruotolo, B. T.; Verbeck, G. F.; Thomson, L. M.; Gillig, K. J.; Russell, D. H. Observation of conserved solution-phase secondary structure in gas-phase tryptic peptides *Journal of the American Chemical Society* **2002**, *124*, 4214-4215.

57.   Ruotolo, B. T.; Verbeck, G. F.; Thomson, L. M.; Woods, A. S.; Gillig, K. J.; Russell, D. H. Distinguishing between phosphorylated and nonphosphorylated peptides with ion mobility-mass spectrometry *Journal of Proteome Research* **2002**, *1*, 303-306.

58.   Ruotolo, B. T.; Gillig, K. J.; Woods, A. S.; Egan, T. F.; Ugarov, M. V.; Schultz, J. A.; Russell, D. H. Analysis of phosphorylated peptides by ion mobility-mass spectrometry *Analytical Chemistry* **2004**, *76*, 6727-6733.

59.   McLean, J. R.; McLean, J. A.; Wu, Z.; Becker, C.; Pérez, L. M.; Pace, C. N.; Scholtz, J. M.; Russell, D. H. Factors that influence helical preferences for singly-charged gas-phase peptide ions: the effects of multiple charge-carrying sites *Journal of Physical Chemistry* **in press**.

60.   Stearns, J. A.; Mercier, S.; Seaiby, C.; Guidi, M.; Boyarkin, O. V.; Rizzo, T. R. Conformation-specific spectroscopy and photodissociation of cold, protonated tyrosine and phenylalanine *Journal of the American Chemical Society* **2007**, *129*, 11814-11820.

61.   Stearns, J. A.; Guidi, M.; Boyarkin, O. V.; Rizzo, T. R. Conformation-specific infrared and ultraviolet spectroscopy of tyrosine-based protonated dipeptides *Journal of Chemical Physics* **2007**, *127*, 154322-154334.

62.   Stearns, J. A.; Boyarkin, O. V.; Rizzo, T. R. Spectroscopic signatures of gas-phase helices: Ac-Phe-(Ala)(5)-Lys-H+ and Ac-Phe-(Ala)(10)-Lys-H+ *Journal of the American Chemical Society* **2007**, *129*, 13820-13821.

63.   Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. Naked protein conformations: cytochrome *c* in the gas phase *Journal of the American Chemical Society* **1995**, *117*, 10141-10142.

64.   Shelimov, K. B.; Jarrold, M. F. Conformations, unfolding, and refolding of apomyoglobin in vacuum: an activation barrier for gas-phase protein folding *Journal of the American Chemical Society* **1997**, *119*, 2987-2994.

65.   Valentine, S. J.; Plasencia, M. D.; Liu, X.; Krishnan, M.; Naylor, S.; Udseth, H. R.; Smith, R. D.; Clemmer, D. E. Toward plasma proteome profiling with ion mobility-mass dpectrometry *Journal of Proteome Research* **2006**, *5*, 2977-2984.

66. Park, Z. Y.; Russell, D. H. Thermal denaturation: a useful technique in peptide mass mapping *Analytical Chemistry* **2000**, *72*, 2667-2670.

67. Wilkins, M. R.; Lindskog, I.; Gasteiger, E.; Bairoch, A.; Sanchez, J.-C.; Hochstrasser, D. F.; Appel, R. D. Detailed peptide characterization using PEPTIDEMASS - a world-wide-web-accessible tool. *Electrophoresis* **1997**, *18*, 403-408.

68. Gillig, K. J.; Ruotolo, B. T.; Stone, E. G.; Russell, D. H.; Fuhrer, K.; Gonin, M.; Schultz, J. A. Coupling high-pressure MALDI with ion mobility/orthogonal time-of-flight mass spectrometry *Analytical Chemistry* **2000**, *72*, 3965-3971.

69. Mason, E. A.; McDaniel, E. W.; *Transport Properties of Ions in Gases*; Wiley: New York, 1988, pp 1-29.

70. Ruotolo, B. T.; McLean, J. A.; Gillig, K. J.; Russell, D. H. The influence and utility of varying field strength for the separation of tryptic peptides by ion mobility-mass spectrometry *Journal of the American Society of Mass Spectometry* **2005**, *16*, 158-165.

71. Verbeck, G. F.; Ruotolo, B. T.; Gillig, K. J.; Russell, D. H. Resolution equations for high-field ion mobility *Journal of the American Society of Mass Spectometry* **2004**, *15*, 1320-1324.

72. Wyttenbach, T.; von Helden, G.; Bowers, M. T. Gas-phase conformation of biological molecules: bradykinin **1996**, *118*, 8355-8364.

73. Sawyer, H. A.; Marini, J. T.; Stone, E. G.; Ruotolo, B. T.; Gillig, K. J.; Russell, D. H. The structure of gas-phase bradykinin fragment 1-5 (RPPGF) ions: an ion mobility spectrometry and H/D exchange ion-molecule reaction chemistry study *Journal of the American Society for Mass Spectrometry* **2005**, *16*, 893-905.

74. Shvartsburg, A. A.; Jarrold, M. F. An exact hard-spheres scattering model for the mobilities of polyatomic ions *Chemical Physics Letters* **1996**, *261*, 86-91.

75. Ruotolo, B. T.; Gillig, K. J.; Stone, E. G.; Russell, D. H. Peak capacity of ion mobility mass spectrometry: separation of peptides in helium buffer gas *Journal of Chromatography B* **2002**, *782*, 385-392.

76. Hudgins, R. R.; Jarrold, M. F. Helix formation in unsolvated alanine-based peptides: helical monomers and helical dimers *Journal of the American Chemical Society* **1999**, *121*, 3494-3501.

77.   Ruotolo, B. T.; Verbeck, G. F.; Thomson, L. M.; Woods, A. S.; Gillig, K. J.; Russell, D. H. Distinguishing between phosphorylated and nonphophorylated peptides with ion mobility-mass spectrometry *Journal of Proteome Research* **2002**, *1*, 303-306.

78.   Henderson, S. C.; Li, J.; Counterman, A. E.; Clemmer, D. E. Intrinsic size parameters for Val, Ile, Leu, Gln, Thr, Phe, and Trp residues from ion mobility measurements of polyamino acid ions *J. Phys. Chem. B* **1999**, *103*, 8780-8785.

79.   Slaton, J. G.; Russell, W. K.; Russell, D. H. Factors affecting gas-phase peptide ion fragmentation: N-terminal residue, charge carrier, and conformation *Journal of the American Society for Mass Spectrometry* **in press**.

80.   Wilson, S. R.; Cui, W. Applications of simulated annealing to pepetides *Biopolymers* **1990**, *29*, 225-235.

81.   Fernandez-Lima, F. A.; Wei, H.; Gao, Y. Q.; Russell, D. H. On the structure elucidation using ion mobility spectrometry and molecular dynamics *The Journal of Physical Chemistry A* **2009**, *113*, 8221-8234.

82.   Damsbo, M.; Kinnear, B. S.; Hartings, M. R.; Ruhoff, P. T.; Jarrold, M. F.; Ratner, M. A. Application of evolutionary algorithm methods to polypeptide folding: comparison with experimental results for unsolvated Ac-(Ala-Gly-GlY)(5)-LysH(+) *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7215-7222.

83.   Reid, G.; Simpson, R.; O'Hair, R. J. A mass spectrometric and ab initio study of the pathways for dehydration of simple glycine and cysteine-containing peptide [M+H]+ ions *Journal of the American Society for Mass Spectrometry* **1998**, *9*, 945-956.

84.   Carpino, L. A.; Han, G. Y. 9-Fluorenylmethoxycarbonyl function, a new base-sensitive amino-protecting group *Journal of the American Chemical Society* **1970**, *92*, 5748-5749.

85.   Damsbo, M.; Kinnear, B. S.; Hartings, M. R.; Ruhoff, P. T.; Jarrold, M. F.; Ratner, M. A. Application of evolutionary algorithm methods to polypeptide folding: comparison with experimental results for unsolvated Ac-(Ala-Gly-Gly)5-LysH+ *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7215-7222.

86.   Dahl, D. B. In *Bayesian Inference for Gene Expression and Proteomics*; Do, K.-A., Müller, P., Vannucci, M., Ed.; Cambridge University Press: Cambridge, 2006, pp 201-218.

87.   Binder, D. A. Bayesian cluster analysis *Biometrika* **1978**, *65*, 31-38.

88.   Hubert, L.; Arabie, P. Comparing partitions *Journal of Classification* **1985**, *2*, 193-218.

89.   Lee, S.-W.; Kim, H. S.; Beauchamp, J. L. Salt bridge chemistry applied to gas-phase peptide sequencing: selective fragmentation of sodiated gas-phase peptide ions adjacent to aspartic acid residues *Journal of the American Chemical Society* **1998**, *120*, 3188-3195.

90.   Ganesh, S.; Jayakumar, R. Role of N-t-Boc group in helix initiation in a novel tetrapeptide *Journal of Peptide Research* **2002**, *59*, 249-256.

91.   Watson, J. D.; Crick, F. H. C. A structure for deoxyribose nucleic acid *Nature* **1953**, *171*, 737-738.

92.   Wang, A. H. J.; Quigley, G. J.; Kolpak, F. J.; Crawford, J. L.; Van Boom, J. H.; Van der Marel, G.; Rich, A. Molecular structure of a left-handed double helical DNA fragment at atomic resolution *Nature* **1979**, *282*, 680-686.

93.   Rothenburg, S.; Koch-Nolte, F.; Haag, F. DNA methylation and Z-DNA formation as mediators of quantitative differences in the expression of alleles *Immunol Rev* **2001**, *184*, 286-298.

94.   Oh, D.; Kim, Y.; Rich, A. Z-DNA-binding proteins can act as potent effectors of gene expression *in vivo Proc. Natl. Acad. Sci. U.S.A* **2002**, *99*, 16666-16671.

95.   Marky, L. A.; Breslauer, K. J. Calorimetric determination of base-stacking enthalpies in double-helical DNA molecules *Biopolymers* **1982**, *21*, 2185-2194.

96.   Clauwaert J; J., S. Interactions of polynucleotides and thier components. I. dissociation constants of the bases and their derivatives. *Z. Naturforsch. B* **1968**, *23*, 25-30.

97.   Ts'o, P. O. P. In *Basic Principles in Nucleic Acid Chemistry*; Ts'o, P. O. P., Ed.; Academic Press: New York, 1974; Vol. 1, pp 454.

98.   Sponer, J.; Leszczynski, J.; Hobza, P. Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases *Biopolymers* **2001**, *61*, 3-31.

99.   Rodgers, M. T.; Campbell, S.; Marzluff, E. M.; Beauchamp, J. L. Low-energy collision-induced dissociation of deprotonated dinucleotides: determination of the energetically favored dissociation pathways and the relative acidities of the nucleic acid bases *International Journal of Mass Spectrometry and Ion Processes* **1994**, *137*, 121-149.

100.   Rodgers, M. T.; Campbell, S.; Marzluff, E. M.; Beauchamp, J. L. Site-specific protonation directs low-energy dissociation pathways of dinucleotides in the gas phase *International Journal of Mass Spectrometry and Ion Processes* **1995**, *148*, 1-23.

101.   Greco, F.; Liguori, A.; Sindona, G.; Uccella, N. Gas-phase proton affinity of deoxyribonucleosides and related nucleobases by fast atom bombardment tandem mass spectrometry *Journal of the American Chemical Society* **2002**, *112*, 9092-9096.

102.   Stults, J. T.; Marsters, J. C.; Carr, S. A. Improved electrospray ionization of synthetic oligodeoxynucleotides *Rapid Communications in Mass Spectrometry* **1991**, *5*, 359-363.

103.   Schneider, K.; Chait, B. T. Matrix-assisted laser desorption mass. spectrometry of homopolymer oligodeoxyribonucleotides *Org. Mass Spectrom.* **1993**, *28*, 1352-1361.

104.   Zhu, L.; Parr, G. R.; Fitzgerald, M. C.; Nelson, C. M.; Smith, L. M. Oligodeoxynucleotide fragmentation in MALDI/TOF mass spectrometry using 355-nm radiation *Journal of the American Chemical Society* **2002**, *117*, 6048-6056.

105.   Nordhoff, E.; Karas, M.; Cramer, R.; Hahner, S.; Hillenkamp, F.; Kirpekar, F.; Lezius, A.; Muth, J.; Meier, C.; Engels, J. W. Direct mass spectrometric sequencing of low-picomole amounts of oligodeoxynucleotides with up to 21 bases by matrix-assisted laser desorption/ionization mass spectrometry *Journal of Mass Spectrometry* **1995**, *30*, 99-112.

106.   Spengler, B.; Pan, Y.; Cotter, R. J.; Kan, L.-S. Molecular weight determination of underivatized oligodeoxyribonucleotides by positive-ion matrix-assisted ultraviolet laser-desorption mass spectrometry *Rapid Communications in Mass Spectrometry* **1990**, *4*, 99-102.

107.   Bornsen, K. O.; Schar, M.; Widmer, H. M. Matrix-assisted laser desorption and ionization mass spectrometry and its application in chemistry *Chimia* **1990**, *44*, 412.

108.   Hillenkamp, F.; Karas, M.; Ingendoh, A.; Stahl, B. In *Biological Mass Spectrometry*; Burlingame, A., McCloskey, J. A., Eds.; Elsevier: Amsterdam, 1990, pp 49.

109. T. Huth-Fehre; J. N. Gosine; K. J. Wu; Becker, C. H. Matrix-assisted laser desorption mass spectrometry of oligodeoxythmidylic acids *Rapid Communications in Mass Spectrometry* **1992**, *6*, 209-213.

110. Gary R. Parr; Michael C. Fitzgerald; Smith, L. M. Matrix-assisted laser desorption/ionization mass spectrometry of synthetic oligodeoxyribonucleotides *Rapid Communications in Mass Spectrometry* **1992**, *6*, 369-372.

111. Kuang Jen Wu; Anna Steding; Becker, C. H. Matrix-assisted laser desorption time-of-flight mass spectrometry of oligonucleotides using 3-hydroxypicolinic acid as an ultraviolet-sensitive matrix *Rapid Communications in Mass Spectrometry* **1993**, *7*, 142-146.

112. Taranenko, N. I.; Tang, K.; Allman, S. L.; Ch'ang, L. Y.; Chen, C. H. 3-aminopicolinic aid as a matrix for laser desorption mass spectrometry of biopolymers *Rapid Communications in Mass Spectrometry* **1994**, *8*, 1001-1006.

113. Pieles, U.; Zurcher, W.; Schar, M.; Moser, H. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry: a powerful tool for the mass and sequence analysis of natural and modified oligonucleotides *Nucl. Acids Res.* **1993**, *21*, 3191-3196.

114. Lecchi, P.; Le, H. M. T.; Pannell, L. K. 6-Aza-2-thiothymine: a matrix for MALDI spectra of oligonucleotides *Nucl. Acids Res.* **1995**, *23*, 1276-1277.

115. Nordhoff, E.; Ingendoh, A.; Cramer, R.; Overberg, A.; Stahl, B.; Karas, M.; Hillenkamp, F.; Crain, P. F.; Chait, B. Matrix-assisted laser desorption/ionization mass spectrometry of nucleic acids with wavelengths in the ultraviolet and infrared *Rapid Communications in Mass Spectrometry* **1992**, *6*, 771-776.

116. Nordhoff, E.; Cramer, R.; Karas, M.; Hillenkamp, F.; Kirpekar, F.; Kristiansen, K.; Roepstorff, P. Ion stability of nucleic acids in infrared matrix-assisted laser desorption/ionization mass spectrometry *Nucl. Acids Res.* **1993**, *21*, 3347-3357.

117. Bai, J.; Liu, Y.-H.; Lubman, D. M.; Siemieniak, D.; Chen, C.-H. Matrix-assisted laser desorption/ionization mass sectrometry of restriction enzyme-digested plasmid DNA using an active nafion substrate *Rapid Communications in Mass Spectrometry* **1994**, *8*, 687-691.

118. Liu, Y.-H.; Bai, J.; Zhu, Y.; Liang, X.; Lubman, D. M.; Siemieniak, D.; Venta, P. J. Rapid screening of genetic polymorphisms using buccal cell DNA with detection by matrix-assisted laser desorption/ionization mass spectrometry *Rapid Communications in Mass Spectrometry* **1995**, *9*, 735-743.

119.   Fitzgerald, M. C.; Zhu, L.; Smith, L. M. The analysis of mock DNA sequencing reactions using matrix-assisted laser desorption/ionization mass spectrometry *Rapid Communications in Mass Spectrometry* **1993**, *7*, 895-897.

120.   Lecchi, P.; Pannell, L. K. The detection of intact double-stranded DNA by MALDI *Journal of the American Society for Mass Spectrometry* **1995**, *6*, 972-975.

121.   Kinnear, B. S.; Jarrold, M. F. Helix formation in unsolvated peptides: side chain entropy is not the determining factor *Journal of the American Chemical Society* **2001**, *123*, 7907-7908.

122.   Jarrold, M. F. Peptides and proteins in the vapor phase *Annual Review of Physical Chemistry* **2000**, *51*, 179-207.

123.   Gidden, J.; Bushnell, J. E.; Bowers, M. T. Gas-phase conformations and folding energetics of oligonucleotides: dTG- and dGT *Journal of the American Chemical Society* **2001**, *123*, 5610-5611.

124.   Gidden, J.; Bowers, M. T. Gas-phase conformations of deprotonated trinucleotides (dGTT−, dTGT−, and dTTG−): the question of zwitterion formation *Journal of the American Society for Mass Spectrometry* **2003**, *14*, 161-170.

125.   Dupradeau, F.-Y.; Cezard, C.; Lelong, R.; Stanislawiak, E.; Pecher, J.; Delepine, J. C.; Cieplak, P. R.E.D.D.B.: a database for RESP and ESP atomic charges, and force field libraries *Nucl. Acids Res.* **2008**, *36*, D360-367.

126.   Gao, Y. Q. An integrate-over-temperature approach for enhanced sampling *The Journal of Chemical Physics* **2008**, *128*, 064105-064105.

127.   Gao, Y. Q.; Yang, L. On the enhanced sampling over energy barriers in molecular dynamics simulations *The Journal of Chemical Physics* **2006**, *125*, 114103-114105.

128.   Yang, L.; Grubb, M. P.; Gao, Y. Q. Application of the accelerated molecular dynamics simulations to the folding of a small protein *The Journal of Chemical Physics* **2007**, *126*, 125102-125107.

129.   Wang, J.; Cieplak, P.; Kollman, P. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry* **2000**, *21*, 1049-1074.

130. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E. e. a.*,* C.02 ed.; Gaussian, Inc.: Wallingford, CT, 2004.

131. NIH research resource center for the development of multiscale modeling tools for structural biology (MMTSB) - http://blue11.bch.msu.edu/mmtsb/cluster.pl. Accessed on Jul., 2009.

132. Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. Mobilities of carbon cluster Ions: critical importance of the molecular attractive potential *Journal of Physical Chemistry* **1998**, *108*, 2416-2423.

133. Koomen, J. M.; Russell, W. K.; Hettick, J. M.; Russell, D. H. Improvement of resolution, mass accuracy, and reproducibility in reflected mode DE-MALDI-TOF analysis of DNA using fast evaporation-overlayer sample preparations *Analytical Chemistry* **2000**, *72*, 3860-3866.

134. Ding, J.; Anderegg, R. J. Specific and nonspecific dimer formation in the electrospray ionization mass spectrometry of oligonucleotides *Journal of the American Society for Mass Spectrometry* **1995**, *6*, 159-164.

135. Tao, L.; Dahl, D. B.; Perez, L. M.; Russell, D. H. The contributions of molecular framework to IMS collision cross-sections of gas-phase peptide ions *Journal of the American Society for Mass Spectrometry* **2009**, *20*, 1593-1602.

136. Figueroa, I. D.; Russell, D. H. Matrix-assisted laser desorption ionization hydrogen/deuterium exchange studies to probe peptide conformational changes *Journal of the American Chemical Society for Mass Spectrometry* **1999**, *10*, 719-731.

137. Buckley, P.; Edison, A. S.; Kemple, M. D.; Prendergast, F. G. 13Ca - NMR assignments of melittin in methanol and chemical shift correlations with secondary structure *Journal of Biomolecular NMR* **1993**, *3*, 639-652.

138. Terwilliger, T. C.; Eisenberg, D. The structure of melittin. II. interpretation of the structure *Journal of Biological Chemistry* **1982**, *257*, 6016-6022.

139. Anderegg, R. J.; Wagner, D. S.; Stevenson, C. L.; Borchardt, R. T. The mass spectrometry of helical unfolding in peptides *Journal of the American Society for Mass Spectrometry* **1994**, *5*, 425-433.

APPENDIX A

THE ION-NEUTRAL COLLISION CROSS-SECTIONS OF 607 PROTEOLYTIC

PEPTIDE IONS MEASURED BY MALDI-IM-TOFMS

Enzyme	Protein	Position	Peptide Sequence	m/z	Ω (Å²)	σ	% Dev.
T	ALBU_BOVIN	229-232	FGER	508.25	147	2	-2.93%
T	ENO1_YEAST	5-14	VYAR	508.29	154	1	1.97%
C	ALBU_BOVIN	147-150	CDEF	513.17	152	2	-0.09%
C	HBB_HORSE	42-45	FDSF	515.21	144	2	-5.63%
P	OVAL_CHICK	84-87	RDIL	516.31	143	1	-6.44%
T	LYSC_CHICK	87-91	TPGSR	517.27	143	3	-6.32%
C	ALBU_BOVIN	91-94	HTLF	517.28	149	1	-2.27%
T	CASB_BOVIN	44-47	KIEK	517.33	154	1	0.88%
C	OVAL_CHICK	145-148	INSW	519.26	155	1	0.89%
P	OVAL_CHICK	66-70	GDSIE	520.22	154	2	0.49%
C	HBA_HORSE	126-129	DKFL	522.29	156	2	1.18%
P	ALDOA_RABIT	82-85	TLYQ	524.27	150	2	-2.45%
C	ALDOA_RABIT	59-62	RQLL	529.35	162	1	4.49%
C	HBB_HORSE	115-118	ARHF	530.28	149	1	-4.23%
P	HBB_HORSE	131-135	QKVVA	544.35	147	1	-6.97%
C	OVAL_CHICK	9-12	EFCF	545.21	156	3	-1.46%
C	ALBU_BOVIN	230-234	GERAL	545.30	164	1	3.35%
T	ALBU_BOVIN	101-105	VASLR	545.34	159	2	0.51%
T	ENO1_YEAST	392-396	TGQIK	546.32	160	1	0.85%
C	MYG_HORSE	30-33	IRLF	548.36	173	1	8.58%
C	OVAL_CHICK	294-298	TSVLM	550.29	161	2	1.18%
C	ALBU_BOVIN	431-434	IVRY	550.33	170	1	6.47%
C	HBB_HORSE	38-41	TQRF	551.29	159	3	-0.03%
C	ALBU_BOVIN	158-161	WGKY	553.28	155	3	-2.82%
T	ENO1_YEAST	120-125	AAAAEK	560.30	163	1	0.84%
C	ALBU_BOVIN	521-525	VPKAF	561.34	162	3	0.41%
C	ALBU_BOVIN	47-51	VLIAF	562.36	169	1	4.75%
C	OVAL_CHICK	218-222	RVASM	563.30	159	2	-1.68%
P	ALDOA_RABIT	324-327	QEEY	568.22	157	2	-3.55%
C	OVAL_CHICK	288-291	EEKY	568.26	161	2	-0.87%
C	ALBU_BOVIN	105-108	RETY	568.27	163	1	-0.17%
C	ALBU_BOVIN	422-426	GEYGF	572.24	156	1	-4.54%
T	ENO1_YEAST	397-402	TGAPAR	572.32	163	2	-0.61%
T	ALBU_BOVIN	219-222	QRLR	572.36	160	3	-1.98%
C	LYSC_CHICK	22-26	GRCEL	577.28	160	3	-2.51%
C	OVAL_CHICK	213-217	QIGLF	577.33	166	2	0.73%
C	ALDOA_RABIT	58-61	YRQL	579.32	171	1	3.69%
T	OVAL_CHICK	123-126	ELYR	580.31	167	4	1.18%
P/C	HBA_HORSE	30-33	ERMF	582.27	159	7	-4.03%
P	CYC_HORSE	44-48	PGFTY	584.27	149	2	-10.19%
T	ALDOA_RABIT	22-27	IVAPGK	584.38	168	6	1.10%
C	ALDOA_RABIT	328-332	VKRAL	586.40	176	2	5.83%

Enzyme	Protein	Position	Peptide Sequence	m/z	Ω (Å^2)	σ	% Dev.
P	ALDOA_RABIT	54-57	NRRF	592.33	153	1	-8.75%
C	OVAL_CHICK	212-216	YQIGL	593.33	171	1	2.05%
C	LYSC_CHICK	39-43	RGYSL	595.32	169	2	0.48%
C	ALBU_BOVIN	365-370	AVSVLL	601.39	176	2	4.09%
C	CASB_BOVIN	104-108	QPEVM	603.28	172	2	1.30%
C	HBB_HORSE	86-91	AALSEL	603.33	167	2	-1.67%
T	OVAL_CHICK	51-55	TQINK	603.35	175	5	3.17%
C	HBA_HORSE	87-91	HAHKL	605.35	171	2	0.83%
T	LYSC_CHICK	19-23	KVFGR	606.37	173	3	1.84%
C	LYSC_CHICK	31-35	KRHGL	610.38	168	2	-1.86%
T	ALDOA_RABIT	134-138	CAQYK	612.28	170	1	-0.87%
C	HBB_HORSE	92-96	HCDKL	615.29	171	2	-0.38%
C	CYC_HORSE	60-64	KEETL	619.33	175	2	1.07%
C	ALBU_BOVIN	56-60	QQCPF	622.27	186	5	7.48%
C	ALBU_BOVIN	479-483	ILNRL	628.41	185	3	5.90%
T	ALDOA_RABIT	101-107	GGVVGIK	629.40	178	3	1.96%
T	MYG_HORSE	140-145	NDIAAK	631.34	179	2	2.19%
T	OVAL_CHICK	182-186	GLWEK	632.34	178	4	1.88%
C	OVAL_CHICK	268-273	TSSNVM	638.28	178	2	1.25%
C	LYSC_CHICK	47-52	VCAAKF	638.33	186	1	5.42%
C	ALDOA_RABIT	59-63	RQLLL	642.43	192	2	8.50%
C	MYG_HORSE	70-76	TALGGIL	644.40	188	1	5.99%
C	ALBU_BOVIN	209-213	REKVL	644.41	182	3	2.57%
T	CASB_BOVIN	115-120	EAMAPK	646.32	182	3	2.16%
P	OVAL_CHICK	100-105	SLASRL	646.39	178	1	0.31%
C	CASB_BOVIN	68-73	AQTQSL	647.34	184	2	3.57%
T	OVAL_CHICK	280-284	VYLPR	647.39	193	3	8.20%
C	ALBU_BOVIN	425-430	GFQNAL	649.33	176	3	-1.39%
T	ALBU_BOVIN	223-228	CASIQK	649.33	182	3	2.20%
C	ALBU_BOVIN	526-530	DEKLF	651.33	180	1	0.79%
P/C	HBA_HORSE	137-141	TSKYR	654.36	176	4	-2.03%
T	ENO1_YEAST	234-240	AAGHDGK	655.32	180	3	0.05%
C	HBB_HORSE	142-146	AHKYH	655.33	180	1	0.04%
P	CYC_HORSE	63-67	TLMEY	656.30	183	1	1.94%
C	ALBU_BOVIN	180-184	YANKY	658.32	183	1	1.89%
T	ENO1_YEAST	79-84	ANIDVK	659.37	186	1	3.40%
C	ALDOA_RABIT	80-84	HETLY	662.31	183	1	1.42%
T	MYG_HORSE	57-62	ASEDLK	662.34	181	3	-0.03%
C	MYG_HORSE	56-61	KASEDL	662.34	186	2	2.71%
P	OVAL_CHICK	175-180	VNAIVF	662.39	179	1	-1.12%
C	HBB_HORSE	33-37	VVYPW	663.35	185	3	2.00%
C	CASB_BOVIN	135-140	TESQSL	664.31	186	2	2.39%

Enzyme	Protein	Position	Peptide Sequence	m/z	Ω (Å^2)	σ	% Dev.
C	ALDOA_RABIT	300-305	SYGRAL	666.36	193	1	6.49%
C	ALBU_BOVIN	158-162	WGKYL	666.36	185	3	1.70%
C	ALDOA_RABIT	337-342	ACQGKY	669.30	196	2	7.42%
P	CASB_BOVIN	16-20	RELEE	675.33	166	1	-9.69%
T	MYG_HORSE	43-47	FDKFK	684.37	192	1	3.70%
C	ALBU_BOVIN	252-257	VEVTKL	688.42	198	3	6.82%
C	ALDOA_RABIT	204-209	VTEKVL	688.42	196	2	5.77%
T	ALBU_BOVIN	236-241	AWSVAR	689.37	189	2	1.48%
C	CASB_BOVIN	149-154	HLPLPL	689.43	195	1	4.80%
C	CYC_HORSE	99-104	KKATNE	690.38	180	1	-3.46%
P	ALDOA_RABIT	58-62	YRQLL	692.41	185	2	-0.95%
C	HBA_HORSE	30-34	ERMFL	695.35	194	2	3.71%
C	ALDOA_RABIT	199-203	KRCQY	697.35	196	2	4.76%
C	HBB_HORSE	38-42	TQRFF	698.36	193	1	2.89%
C	HBA_HORSE	74-80	DDLPGAL	700.35	192	1	2.36%
C	ALDOA_RABIT	358-363	ISNHAY	704.34	192	2	1.54%
T	MYG_HORSE	51-56	TEAEMK	708.32	196	1	3.61%
C	MYG_HORSE	50-55	KTEAEM	708.32	195	0	3.01%
C	HBA_HORSE	67-73	TLAVGHL	710.42	195	2	2.50%
T	ALBU_BOVIN	29-34	SEIAHR	712.37	188	1	-1.00%
C	LYSC_CHICK	142-147	IRGCRL	717.42	194	1	1.28%
C	OVAL_CHICK	119-124	QCVKEL	719.38	192	2	0.24%
C	MYG_HORSE	124-131	GADAQGAM	720.30	204	1	6.46%
T	ALDOA_RABIT	140-146	DGADFAK	723.33	196	5	2.08%
T	ENO1_YEAST	60-66	GVLHAVK	723.45	205	2	6.46%
C	ALBU_BOVIN	515-520	TPDETY	725.30	190	2	-1.46%
T	ENO1_YEAST	9-14	SVYDSR	726.34	192	3	-0.61%
C	ALDOA_RABIT	292-297	LKPWAL	727.45	203	1	4.98%
C	ALBU_BOVIN	49-54	IAFSQY	728.36	194	3	0.22%
P	CASB_BOVIN	214-220	GPVRGPF	729.40	190	2	-1.67%
P	ALBU_BOVIN	370-375	LRLAKE	729.46	195	1	0.60%
P	ALBU_BOVIN	237-242	WSVARL	731.42	190	1	-2.10%
C	ALBU_BOVIN	55-60	LQQCPF	735.35	198	3	1.78%
T	MYG_HORSE	97-102	HKIPIK	735.49	209	1	7.64%
C	ALBU_BOVIN	61-66	DEHVKL	740.39	199	4	1.82%
C	OVAL_CHICK	212-217	YQIGLF	740.40	198	1	1.37%
C	OVAL_CHICK	359-364	RADHPF	742.36	191	1	-2.53%
T	CASB_BOVIN	218-224	GPFPIIV	742.45	203	2	3.84%
T	ENO1_YEAST	330-336	IATAIEK	745.45	210	2	6.74%
T	CASB_BOVIN	123-128	EMPFPK	748.37	201	1	2.11%
T	MYG_HORSE	134-139	ALELFR*	748.44	215	0	9.32%
C	CYC_HORSE	60-65	KEETLM	750.37	198	5	0.48%

Enzyme	Protein	Position	Peptide Sequence	m/z	Ω (Å^2)	σ	% Dev.
T	ENO1_YEAST	409-414	LNQLLR	756.47	215	1	8.54%
P	OVAL_CHICK	84-89	RDILNQ	758.42	184	1	-7.33%
T	CYC_HORSE	8-13	KIFVQK	762.49	225	3	12.93%
P/C	OVA_CHICK	379-385	FGRCVSP	765.37	195	4	-2.61%
C	ALDOA_RABIT	217-222	SDHHIY	771.34	205	1	1.84%
C	MYG_HORSE	147-153	KELGFQG	778.41	207	1	2.16%
C	CYC_HORSE	69-74	ENPKKY	778.41	197	1	-2.84%
C	ALBU_BOVIN	371-376	RLAKEY	779.44	205	5	1.42%
T	CYC_HORSE	80-86	MIFAGIK*	779.45	223	4	10.24%
T	OVAL_CHICK	105-110	LYAEER	780.39	207	3	2.11%
C	ALDOA_RABIT	138-144	KKDGADF	780.39	214	1	5.69%
T	CASB_BOVIN	185-191	VLPVPQK	780.50	208	8	2.84%
T	ENO1_YEAST	132-138	HLADLSK	783.44	212	2	4.19%
C	ALBU_BOVIN	140-146	KPDPNTL	784.42	206	1	1.40%
P/C	HBB_HORSE	104-110	RLLGNVL	784.50	214	6	5.17%
P	CYC_HORSE	62-67	ETLMEY	785.34	202	1	-0.95%
C	CASB_BOVIN	172-178	FPPQSVL	787.43	210	2	2.93%
C	CASB_BOVIN	16-21	RELEEL	788.41	205	1	0.52%
P	HBB_HORSE	15-20	WDKVNE	790.37	202	1	-1.25%
P	OVAL_CHICK	109-114	ERYPIL	790.45	200	1	-2.35%
T	OVAL_CHICK	285-290	MKMEEK	795.37	213	3	3.57%
C	HBA_HORSE	118-125	TPAVHASL	795.44	207	2	0.75%
C	OVAL_CHICK	235-242	ASGTMSML	797.35	205	1	-0.27%
C	HBB_HORSE	42-48	FDSFGDL	800.35	212	2	2.98%
P	OVAL_CHICK	138-144	ADQAREL	802.41	198	2	-4.11%
C	LYSC_CHICK	75-80	QINSRW	803.42	202	2	-2.14%
C	CYC_HORSE	98-104	LKKATNE	803.46	213	2	2.73%
C	HBA_HORSE	129-136	LSSVSTVL	805.47	214	3	3.29%
P	ALDOA_RABIT	58-63	YRQLLL	805.49	214	2	3.15%
P	MYG_HORSE	138-144	FRNDIAA	806.42	206	1	-0.58%
T	CYC_HORSE	73-79	KYIPGTK	806.48	225	4	8.58%
T	ENO1_YEAST	178-184	TFAEALR	807.44	220	5	6.12%
C	ALBU_BOVIN	343-349	QEAKDAF	808.38	204	4	-1.64%
C	ALDOA_RABIT	79-84	FHETLY	809.38	218	2	4.74%
C	OVAL_CHICK	100-106	SLASRLY	809.45	215	1	3.49%
P/C	HBB_HORSE	107-114	GNVLVVVL	812.52	203	6	-2.68%
C	MYG_HORSE	41-46	EKFDKF	813.41	214	2	2.67%
C	HBA_HORSE	99-105	KLLSHCL	813.47	208	2	-0.17%
C	ALBU_BOVIN	179-184	YYANKY	821.38	217	2	3.41%
C	MYG_HORSE	132-138	TKALELF	821.48	227	2	7.99%
T	CASB_BOVIN	192-198	AVPYPQR	830.45	218	1	2.80%
C	OVAL_CHICK	118-124	LQCVKEL	832.46	217	2	2.33%

Enzyme	Protein	Position	Peptide Sequence	m/z	Ω (Å^2)	σ	% Dev.
P	ALBU_BOVIN	95-102	GDELCKVA	834.40	208	2	-2.31%
P	HBB_HORSE	36-41	PWTQRF	834.43	211	1	-0.86%
C	ALDOA_RABIT	321-327	KAAQEEY	838.39	215	2	0.95%
T	ALBU_BOVIN	483-489	LCVLHEK	841.46	221	2	3.24%
C	HBA_HORSE	92-98	RVDPVNF	846.45	214	2	-0.44%
P	HBA_HORSE	92-98	RVDPVNF	846.45	206	2	-3.83%
P	ALDOA_RABIT	357-363	FISNHAY	851.40	216	2	0.06%
C	OVAL_CHICK	299-306	AMGITDVF	853.41	219	2	1.68%
P/C	OVA_CHICK	359-365	RADHPFL	855.45	214	5	-0.92%
C	OVAL_CHICK	228-234	KILELPF	859.53	223	1	2.84%
C	ALBU_BOVIN	151-157	KADEKKF	865.48	223	2	2.21%
C	ALBU_BOVIN	568-574	KTVMENF	868.42	224	2	2.53%
T	ALDOA_RABIT	201-207	CQYVTEK	870.40	221	4	1.16%
P	HBB_HORSE	23-31	VGGEALGRL	871.50	218	1	-0.28%
T	CASB_BOVIN	113-120	VKEAMAPK*	873.49	237	6	7.98%
T	LYSC_CHICK	33-39	HGLDNYR	874.42	219	2	-0.27%
P	OVAL_CHICK	196-202	MPFRVTE	879.44	212	1	-3.67%
C	MYG_HORSE	62-69	KKHGTVVL	881.56	229	1	3.68%
C	OVAL_CHICK	119-125	QCVKELY	882.44	229	3	3.56%
C	CASB_BOVIN	61-67	QDKIHPF	884.46	228	2	3.22%
T	OVAL_CHICK	278-284	IKVYLPR	888.57	238	3	7.05%
C	ALBU_BOVIN	505-511	VNRRPCF	891.46	226	2	1.52%
C	CYC_HORSE	68-74	LENPKKY	891.49	226	2	1.60%
P	HBA_HORSE	30-36	ERMFLGF	899.44	220	1	-1.59%
C	MYG_HORSE	33-40	FTGHPETL	901.44	235	1	4.61%
P	LYSC_CHICK	141-147	WIRGCRL	903.50	219	2	-2.29%
C	OVAL_CHICK	378-385	FFGRCVSP	912.44	235	1	4.04%
C	ALBU_BOVIN	181-188	ANKYNGVF	912.46	232	2	2.55%
P	ALBU_BOVIN	165-171	IARRHPY	912.52	224	1	-0.96%
C	HBA_HORSE	35-42	GFPTTKTY	914.46	232	2	2.31%
C	OVAL_CHICK	30-38	CPIAIMSAL	918.48	236	2	4.04%
C	LYSC_CHICK	22-30	GRCELAAAM	921.43	227	2	-0.43%
T	ALBU_BOVIN	249-256	AEFVEVTK	922.49	231	3	1.43%
C	ALBU_BOVIN	373-380	AKEYEATL	924.47	228	6	-0.18%
T	ALBU_BOVIN	161-167	YLYEIAR	927.49	244	1	6.84%
T	LYSC_CHICK	80-86	WWCNDGR	936.38	225	2	-2.13%
T	ALDOA_RABIT	322-329	AAQEEYVK	937.46	238	3	3.19%
T	ALDOA_RABIT	14-21	ELSDIAHR*	940.48	233	1	0.75%
C	HBB_HORSE	111-118	VVVLARHF	940.57	247	1	7.09%
T	MYG_HORSE	146-153	YKELGFQG*	941.47	244	2	5.69%
C	ALBU_BOVIN	358-364	SRRHPEY	944.47	217	2	-6.22%
C	MYG_HORSE	139-146	RNDIAAKY	950.51	235	1	0.96%

Enzyme	Protein	Position	Peptide Sequence	m/z	Ω (Å²)	σ	% Dev.
P	HBB_HORSE	139-146	NALAHKYH	953.50	241	1	3.25%
C	LYSC_CHICK	27-35	AAAMKRHGL	954.53	232	1	-0.81%
C	HBB_HORSE	46-55	GDLSNPGAVM	960.45	234	2	-0.16%
C	OVAL_CHICK	175-183	VNAIVFKGL	960.59	239	2	1.80%
P	OVAL_CHICK	112-119	PILPEYLQ	972.54	213	1	-9.78%
C	ALBU_BOVIN	36-43	KDLGEEHF	974.46	234	3	-1.19%
C	HBB_HORSE	115-122	ARHFGKDF	977.50	241	2	1.63%
P	CYC_HORSE	75-83	IPGTKMIFA	977.55	233	1	-1.96%
C	OVAL_CHICK	181-188	KGLWEKAF	978.54	244	2	2.54%
C	MYG_HORSE	107-115	ISDAIIHVL	980.58	256	4	7.65%
C	LYSC_CHICK	36-43	DNYRGYSL	987.45	235	1	-1.86%
P	OVAL_CHICK	135-143	QTAADQARE	989.46	214	1	-10.52%
C	LYSC_CHICK	75-81	QINSRWW	989.50	235	1	-1.91%
C	ALBU_BOVIN	411-418	IKQNCDQF	995.46	242	4	0.73%
P	LYSC_CHICK	44-52	GNWVCAAKF	995.48	232	2	-3.62%
C	ALBU_BOVIN	422-430	GEYGFQNAL	998.46	243	3	0.96%
C	LYSC_CHICK	31-38	KRHGLDNY	1002.51	236	2	-2.47%
C	OVAL_CHICK	359-366	RADHPFLF	1002.52	248	2	2.42%
P	OVAL_CHICK	358-365	FRADHPFL	1002.52	245	0	1.42%
C	HBA_HORSE	15-24	SKVGGHAGEY	1004.48	249	1	2.99%
P	OVAL_CHICK	196-203	MPFRVTEQ	1007.50	236	1	-2.83%
P	ALDOA_RABIT	217-224	SDHHIYLE	1013.47	248	1	1.73%
T	CASB_BOVIN	121-128	HKEMPFPK*	1013.52	249	2	2.27%
P/C	CYC_HORSE	37-46	GRKTGQAPGF	1018.54	241	5	-1.33%
P	LYSC_CHICK	19-27	KVFGRCELA	1022.55	242	2	-1.33%
C	HBA_HORSE	25-33	GAEALERMF	1023.49	252	1	2.84%
T	ALBU_BOVIN	499-507	CCTESLVNR	1024.46	252	3	2.82%
T	ALDOA_RABIT	200-207	RCQYVTEK	1026.50	253	1	2.90%
C	HBA_HORSE	34-42	LGFPTTKTY	1027.55	251	2	2.08%
T	LYSC_CHICK	32-39	RHGLDNYR	1030.52	243	2	-1.50%
P	ALBU_BOVIN	164-171	EIARRHPY	1041.56	248	1	-0.29%
C	OVAL_CHICK	243-252	VLLPDEVSGL	1041.58	252	3	1.64%
C	CYC_HORSE	60-67	KEETLMEY	1042.48	246	2	-0.85%
T	ALDOA_RABIT	60-68	QLLLTADDR	1044.57	261	2	4.88%
T	LYSC_CHICK	135-143	GTDVQAWIR*	1045.54	256	2	2.70%
P	OVAL_CHICK	275-282	ERKIKVYL	1048.65	250	0	0.41%
C	ALDOA_RABIT	165-173	ENANVLARY	1049.54	262	2	4.89%
P/C	ALDOA_RABIT	152-161	KIGEHTPSAL	1052.57	249	10	-0.61%
P/C	ALDOA_RABIT	333-342	ANSLACQGKY	1054.50	247	12	-1.51%
C	CASB_BOVIN	22-31	NVPGEIVESL	1056.56	259	2	3.13%
P	CASB_BOVIN	22-31	NVPGEIVESL	1056.56	246	2	-1.92%
P	ALBU_BOVIN	165-172	IARRHPYF	1059.58	251	1	-0.14%

Enzyme	Protein	Position	Peptide Sequence	m/z	Ω (Å^2)	σ	% Dev.
C	OVAL_CHICK	274-281	EERKIKVY	1064.61	255	2	1.16%
T	ALDOA_RABIT	13-21	KELSDIAHR*	1068.58	254	2	0.58%
C	HBB_HORSE	131-141	QKVVAGVANAL	1069.64	260	2	2.73%
P/C	ALBU_BOVIN	243-251	SQKFPKAEF	1081.57	253	4	-0.56%
C	ALDOA_RABIT	214-222	KALSDHHIY	1083.56	265	2	3.68%
C	MYG_HORSE	47-55	KHLKTEAEM	1086.56	266	2	4.07%
C	LYSC_CHICK	72-80	GILQINSRW	1086.61	268	2	4.82%
P/C	HBB_HORSE	4--14	SGEEKAAVLAL	1087.60	257	5	0.34%
T	HBA_HORSE	91-99	LRVDPVNFK	1087.63	280	2	9.38%
P/C	HBA_HORSE	92-100	RVDPVNFKL	1087.63	257	4	0.50%
P	LYSC_CHICK	19-28	KVFGRCELAA	1093.58	255	2	-0.61%
P	MYG_HORSE	138-146	FRNDIAAKY	1097.57	261	1	1.30%
P	ALDOA_RABIT	296-305	ALTFSYGRAL	1098.59	263	1	2.13%
C	OVAL_CHICK	135-144	QTAADQAREL	1102.55	253	2	-1.97%
P	ALBU_BOVIN	357-364	YSRRHPEY	1107.53	244	2	-5.77%
C	OVAL_CHICK	218-227	RVASMASEKM	1109.54	261	1	0.50%
C	HBB_HORSE	69-78	HSFGEGVHHL	1119.53	262	2	0.32%
C	OVAL_CHICK	30-40	CPIAIMSALAM	1120.56	263	3	0.84%
C	CYC_HORSE	1-10	GDVEKGKKIF	1120.64	262	3	0.09%
T	HBB_HORSE	96-104	LHVDPENFR	1126.56	273	2	4.14%
P	HBB_HORSE	21-31	EEVGGEALGRL	1129.58	256	1	-2.48%
P	ALDOA_RABIT	86-95	KADDGRPFPQ	1130.56	254	2	-3.49%
P	OVAL_CHICK	357-365	EFRADHPFL	1131.56	258	1	-1.88%
P	ALDOA_RABIT	327-336	YVKRALANSL	1134.66	267	1	1.19%
C	ALDOA_RABIT	300-310	SYGRALQASAL	1136.61	267	2	1.02%
T	CASB_BOVIN	113-122	VKEAMAPKHK	1138.64	265	1	0.29%
C	ALDOA_RABIT	128-137	DGLSERCAQY	1141.49	267	2	0.87%
T	HBB_HORSE	133-144	VVAGVANALAHK	1149.67	292	2	9.84%
C	OVAL_CHICK	162-172	QPSSVDSQTAM	1150.50	270	1	1.27%
P/C	CASB_BOVIN	214-224	GPVRGPFPIIV	1151.69	276	6	3.65%
P	OVAL_CHICK	106-114	YAEERYPIL	1153.59	266	1	-0.30%
C	MYG_HORSE	34-43	TGHPETLEKF	1158.58	275	3	2.96%
T	ALBU_BOVIN	66-75	LVNELTEFAK	1163.63	279	3	3.83%
P	LYSC_CHICK	19-29	KVFGRCELAAA	1164.62	268	2	-0.32%
P	OVAL_CHICK	125-134	YRGGLEPINF	1165.60	266	1	-1.07%
P/C	ALDOA_RABIT	138-147	KKDGADFAKW	1165.60	274	1	1.88%
C	CYC_HORSE	36-46	FGRKTGQAPGF	1165.61	277	2	2.96%
T	CYC_HORSE	28-38	TGPNLHGLFGR*	1168.62	286	3	6.13%
P	OVAL_CHICK	131-141	PINFQTAADQA	1175.57	267	2	-1.06%
C	MYG_HORSE	77-86	KKKGHHEAEL	1176.65	274	3	1.44%
P	OVAL_CHICK	60-70	DKLPGFGDSIE	1177.57	269	1	-0.50%
C	HBA_HORSE	69-80	AVGHLDDLPGAL	1177.62	281	3	3.94%

Enzyme	Protein	Position	Peptide Sequence	m/z	Ω (Å^2)	σ	% Dev.
P	OVAL_CHICK	265-274	TEWTSSNVME	1183.49	271	2	-0.27%
C	ALBU_BOVIN	333-342	AEDKDVCKNY	1184.53	266	4	-2.20%
T	OVAL_CHICK	360-369	ADHPFLFCIK	1190.60	293	4	7.57%
C	ALBU_BOVIN	371-380	RLAKEYEATL	1193.65	284	3	3.90%
C	HBB_HORSE	104-114	RLLGNVLVVVL	1194.79	304	0	11.32%
C	LYSC_CHICK	42-52	SLGNWVCAAKF	1195.59	265	2	-2.98%
P	HBB_HORSE	33-41	VVYPWTQRF	1195.63	273	1	0.01%
C	ALBU_BOVIN	61-70	DEHVKLVNEL	1195.63	284	3	3.74%
C	HBA_HORSE	126-136	DKFLSSVSTVL	1195.66	304	2	11.21%
C	CASB_BOVIN	93-103	TQTPVVVPPFL	1197.69	278	2	1.65%
P	HBB_HORSE	86-96	AALSELHCDKL	1199.61	278	2	1.49%
C	HBA_HORSE	92-101	RVDPVNFKLL	1200.71	278	2	1.21%
C	CASB_BOVIN	149-158	HLPLPLLQSW	1203.69	276	1	0.35%
P	OVAL_CHICK	16-25	KELKVHHANE	1204.64	267	1	-3.02%
C	ALBU_BOVIN	343-353	QEAKDAFLGSF	1212.59	277	2	0.25%
C	OVAL_CHICK	1-12	GSIGAASMEFCF	1219.51	277	2	-0.15%
C	MYG_HORSE	1-11	GLSDGEWQQVL	1231.60	286	2	2.62%
C	HBA_HORSE	81-91	SNLSDLHAHKL	1234.65	282	2	0.91%
P	ALBU_BOVIN	565-574	EQLKTVMENF	1238.61	263	2	-6.00%
P	HBA_HORSE	37-46	PTTKTYFPHF	1238.62	283	1	1.16%
P	HBB_HORSE	63-73	HGKKVLHSFGE	1238.66	284	1	1.27%
C	HBB_HORSE	1-12	VQLSGEEKAAVL	1243.69	288	2	2.42%
C	CYC_HORSE	49-59	TDANKNKGITW	1247.64	273	3	-3.00%
T	ALBU_BOVIN	35-44	FKDLGEEHFK*	1249.62	290	2	2.77%
T	HBB_HORSE	105-116	LLGNVLVVVLAR*	1265.83	343	2	20.61%
T	LYSC_CHICK	40-51	GYSLGNWVCAAK	1268.61	280	2	-1.72%
T	MYG_HORSE	32-42	LFTGHPETLEK*	1271.66	298	1	4.34%
C	ALBU_BOVIN	36-46	KDLGEEHFKGL	1272.66	287	4	0.55%
C	LYSC_CHICK	72-81	GILQINSRWW	1272.68	288	2	0.85%
T	HBB_HORSE	31-40	LLVVYPWTQR*	1274.73	314	2	10.03%
T	LYSC_CHICK	133-143	CKGTDVQAWIR	1276.65	289	2	1.10%
P	OVAL_CHICK	124-134	LYRGGLEPINF	1278.68	280	1	-2.21%
C	CYC_HORSE	37-48	GRKTGQAPGFTY	1282.65	287	3	-0.06%
T	ALBU_BOVIN	361-371	HPEYAVSVLLR	1283.71	309	3	7.64%
P	MYG_HORSE	30-40	IRLFTGHPETL	1283.71	297	1	3.38%
P	CASB_BOVIN	16-26	RELEELNVPGE	1284.64	281	1	-2.23%
C	LYSC_CHICK	82-93	CNDGRTPGSRNL	1289.60	283	3	-1.81%
C	LYSC_CHICK	19-30	KVFGRCELAAAM	1295.66	287	3	-0.61%
T	CYC_HORSE	28-39	TGPNLHGLFGRK*	1296.72	303	4	4.89%
C	HBA_HORSE	118-129	TPAVHASLDKFL	1298.71	293	2	1.10%
P	OVAL_CHICK	199-209	RVTEQESKPVQ	1300.69	282	1	-2.70%
T	ALBU_BOVIN	402-412	HLVDEPQNLIK*	1305.72	301	1	3.71%

Enzyme	Protein	Position	Peptide Sequence	m/z	Ω (Å²)	σ	% Dev.
P	HBB_HORSE	32-41	LVVYPWTQRF	1308.71	293	1	0.62%
C	HBB_HORSE	43-55	DSFGDLSNPGAVM	1309.57	297	3	1.90%
P	OVAL_CHICK	298-310	MAMGITDVFSSSA	1316.59	276	1	-5.57%
C	ALDOA_RABIT	31-43	AADESTGSIAKRL	1318.70	292	2	-0.06%
C	ALDOA_RABIT	271-283	SGGQSEEEASINL	1320.59	292	2	-0.18%
C	HBB_HORSE	107-118	GNVLVVVLARHF	1323.79	303	1	3.38%
C	OVAL_CHICK	282-291	LPRMKMEEKY	1324.68	303	2	3.10%
C	HBA_HORSE	106-117	LSTLAVHLPNDF	1326.71	301	3	2.44%
C	ALBU_BOVIN	435-446	TRKVPQVSTPTL	1326.77	294	2	-0.05%
P	HBB_HORSE	36-45	PWTQRFFDSF	1330.62	292	1	-0.82%
C	ALBU_BOVIN	287-298	ICDNQDTISSKL	1336.64	290	3	-1.90%
C	MYG_HORSE	136-146	ELFRNDIAAKY	1339.70	303	1	2.44%
C	ALBU_BOVIN	25-35	DTHKSEIAHRF	1340.67	288	4	-2.53%
T	ALDOA_RABIT	1-12	PHSHPALTPEQK*	1341.69	300	1	1.18%
C	HBB_HORSE	33-42	VVYPWTQRFF	1342.69	304	2	2.45%
T	ALDOA_RABIT	87-98	ADDGRPFPQVIK	1342.71	297	1	0.14%
C	ALBU_BOVIN	56-66	QQCPFDEHVKL	1343.64	288	4	-2.70%
P	ALDOA_RABIT	212-222	VYKALSDHHIY	1345.69	302	2	1.69%
T	OVAL_CHICK	370-381	HIATNAVLFFGR*	1345.74	327	2	10.11%
P/C	LYSC_CHICK	27-38	AAAMKRHGLDNY	1346.66	289	2	-2.53%
T	CYC_HORSE	89-99	TEREDLIAYLK	1350.73	320	4	7.66%
P	ALBU_BOVIN	178-188	LYYANKYNGVF	1351.67	291	2	-2.28%
P	MYG_HORSE	138-148	FRNDIAAKYKE	1354.71	300	1	0.80%
C	HBB_HORSE	119-130	GKDFTPELQASY	1355.65	301	3	0.92%
C	MYG_HORSE	104-115	LEFISDAIIHVL	1369.77	325	2	8.20%
C	OVAL_CHICK	19-29	KVHHANENIFY	1371.68	301	2	0.10%
C	ALDOA_RABIT	244-256	SHEEIAMATVTAL	1372.68	308	1	2.36%
C	LYSC_CHICK	31-41	KRHGLDNYRGY	1378.70	281	1	-6.75%
T	MYG_HORSE	64-77	HGTVVLTALGGILK*	1378.84	361	2	19.58%
C	CASB_BOVIN	160-171	HQPHQPLPPTVM	1381.70	313	1	3.43%
T	CASB_BOVIN	206-217	LLYQEPVLGPVR*	1383.80	330	2	8.99%
C	CYC_HORSE	83-94	AGIKKKTEREDL	1387.79	308	5	1.69%
P	CYC_HORSE	83-94	AGIKKKTEREDL	1387.79	306	1	0.99%
C	ALDOA_RABIT	343-357	TPSGQAGAAASESLF	1393.66	305	1	0.25%
P	OVAL_CHICK	184-194	WEKAFKDEDTQ	1396.64	275	2	-9.68%
C	ALDOA_RABIT	284-295	NAINKCPLLKPW	1396.78	311	5	2.06%
C	ALBU_BOVIN	189-201	QECCQAEDKGACL	1397.55	303	4	-0.65%
P	ALBU_BOVIN	151-161	KADEKKFWGKY	1399.74	285	1	-6.56%
C	ALDOA_RABIT	257-269	RRTVPPAVTGVTF	1400.80	309	1	1.20%
C	ALDOA_RABIT	321-332	KAAQEEYVKRAL	1405.78	309	2	0.99%
C	OVAL_CHICK	307-321	SSSANLSGISSAESL	1409.68	310	2	1.09%
T	ENO1_YEAST	105-119	LGANAILGVSLAASR*	1412.82	366	2	19.23%

Enzyme	Protein	Position	Peptide Sequence	m/z	Ω (Å²)	σ	% Dev.
C	ALBU_BOVIN	358-369	SRRHPEYAVSVL	1413.76	306	3	-0.36%
P/C	HBB_HORSE	115-126	ARHFGKDFTPEL	1417.72	309	8	0.40%
P	ALBU_BOVIN	555-566	LLKHKPKATEEQ	1421.81	299	3	-2.93%
C	HBB_HORSE	1-14	VQLSGEEKAAVLAL	1427.81	315	2	1.91%
T	LYSC_CHICK	52-63	FESNFNTQATNR*	1428.65	300	3	-3.08%
C	OVAL_CHICK	199-210	RVTEQESKPVQM	1431.73	303	1	-2.21%
P	HBA_HORSE	87-98	HAHKLRVDPVNF	1432.78	302	2	-2.36%
C	LYSC_CHICK	130-141	RNRCKGTDVQAW	1433.71	295	2	-4.82%
T	CYC_HORSE	26-38	HKTGPNLHGLFGR*	1433.78	337	3	8.91%
T	ALBU_BOVIN	360-371	RHPEYAVSVLLR*	1439.81	321	1	3.25%
C	HBA_HORSE	35-46	GFPTTKTYFPHF	1442.71	314	2	0.96%
C	ALBU_BOVIN	55-66	GERALKAWSVARL	1456.84	318	4	1.46%
P	CASB_BOVIN	211-224	PVLGPVRGPFPIIV	1460.90	324	1	3.16%
C	CASB_BOVIN	49-60	QSEEQQQTEDEL	1463.61	321	2	2.10%
T	CYC_HORSE	40-53	TGQAPGFTYTDANK*	1470.69	317	3	0.60%
C	HBA_HORSE	1-14	VLSAADKTNVKAAW	1473.81	321	3	1.78%
P	CASB_BOVIN	143-155	TDVENLHLPLPLL	1473.83	331	1	4.80%
C	LYSC_CHICK	81-93	WCNDGRTPGSRNL	1475.68	303	2	-4.06%
T	CYC_HORSE	89-100	TEREDLIAYLKK	1478.82	335	4	5.77%
T	ALBU_BOVIN	421-433	LGEYGFQNALIVR*	1479.80	331	1	4.70%
T	ALDOA_RABIT	28-42	GILAADESTGSIAKR	1488.80	315	5	-0.78%
T	ALDOA_RABIT	43-55	LQSIGTENTEENR*	1490.71	306	1	-3.83%
C	OVAL_CHICK	107-118	AEERYPILPEYL	1492.77	327	1	2.76%
T	MYG_HORSE	119-133	HPGDFGADAQGAMTK	1502.67	320	4	-0.04%
T	MYG_HORSE	64-78	HGTVVLTALGGILKK*	1506.94	378	5	17.80%
C	MYG_HORSE	62-76	KKHGTVVLTALGGIL	1506.94	331	1	3.26%
T	ALBU_BOVIN	438-451	VPQVSTPTLVEVSR	1511.84	339	3	5.50%
C	LYSC_CHICK	22-35	GRCELAAAMKRHGL	1512.79	303	3	-5.68%
P	ALBU_BOVIN	151-162	KADEKKFWGKYL	1512.82	315	2	-1.98%
C	MYG_HORSE	77-89	KKKGHHEAELKPL	1514.88	335	2	4.00%
T	HBA_HORSE	17-31	VGGHAGEYGAEALER	1515.72	322	2	0.08%
C	MYG_HORSE	73-86	GGILKKKGHHEAEL	1516.86	334	2	3.88%
C	MYG_HORSE	15-29	GKVEADIAGHGQEVL	1522.79	333	2	3.30%
C	HBB_HORSE	72-85	GEGVHHLDNLKGTF	1523.76	325	2	0.50%
C	ALDOA_RABIT	148-161	RCVLKIGEHTPSAL	1523.84	333	5	3.17%
P	MYG_HORSE	139-151	RNDIAAKYKELGF	1524.82	332	1	2.68%
C	MYG_HORSE	56-69	KASEDLKKHGTVVL	1524.87	343	2	6.24%
C	ALBU_BOVIN	358-370	SRRHPEYAVSVLL	1526.84	327	1	1.12%
P	OVAL_CHICK	186-198	KAFKDEDTQAMPF	1527.71	323	1	-0.23%
C	ALDOA_RABIT	64-78	TADDRVNPCIGGVIL	1542.79	333	5	2.18%
P	HBB_HORSE	115-127	ARHFGKDFTPELQ	1545.78	323	2	-0.95%
C	ALDOA_RABIT	16-30	SDIAHRIVAPGKGIL	1546.91	338	5	3.54%

Enzyme	Protein	Position	Peptide Sequence	m/z	Ω (Å²)	σ	% Dev.
C	MYG_HORSE	34-46	TGHPETLEKFDKF	1548.77	332	2	1.80%
P/C	ALBU_BOVIN	398-410	DKLKHLVDEPQNL	1548.84	328	2	0.39%
T	OVAL_CHICK	187-199	AFKDEDTQAMPFR*	1555.72	330	2	0.88%
C	OVAL_CHICK	199-211	RVTEQESKPVQMM	1562.77	326	3	-0.85%
C	ALDOA_RABIT	314-327	GGKKENLKAAQEEY	1564.80	323	1	-1.90%
P	OVAL_CHICK	13-25	DVFKELKVHHANE	1565.81	319	1	-3.15%
T	ALBU_BOVIN	347-359	DAFLGSFLYEYSR*	1567.74	343	2	4.30%
C	HBB_HORSE	131-145	QKVVAGVANALAHKY	1568.89	340	2	3.24%
T	OVAL_CHICK	187-199	AFKDEDTQAMPFR MSO: 196	1571.72	321	4	-2.81%
P	CYC_HORSE	83-96	AGIKKKTEREDLIA	1571.91	346	0	4.92%
C	ALBU_BOVIN	61-73	DEHVKLVNELTEF	1572.79	334	3	1.19%
P	OVAL_CHICK	11-23	CFDVFKELKVHHA	1572.80	327	2	-0.77%
C	CASB_BOVIN	141-154	TLTDVENLHLPLPL	1574.88	346	3	4.88%
P	OVAL_CHICK	16-28	KELKVHHANENIF	1578.84	328	2	-0.77%
T	OVAL_CHICK	264-276	LTEWTSSNVMEER*	1581.72	321	2	-3.07%
P	CYC_HORSE	22-36	KGGKHKTGPNLHGLF	1590.89	328	1	-1.25%
C	ALDOA_RABIT	306-320	QASALKAWGGKKENL	1600.88	332	3	-0.49%
C	OVAL_CHICK	135-148	QTAADQARELINSW	1602.79	326	1	-2.54%
P	LYSC_CHICK	127-140	VAWRNRCKGTDVQA	1603.81	325	1	-2.76%
T	MYG_HORSE	17-31	VEADIAGHGQEVLIR*	1606.85	340	1	1.50%
T	CYC_HORSE	88-100	KTEREDLIAYLKK	1606.92	351	3	4.83%
C	MYG_HORSE	90-103	AQSHATKHKIPIKY	1621.92	345	2	2.43%
C	MYG_HORSE	116-131	HSKHPGDFGADAQGAM	1625.71	323	2	-4.18%
P	CYC_HORSE	49-62	TDANKNKGITWKEE	1633.82	316	1	-6.77%
T	CYC_HORSE	9-22	IFVQKCAQCHTVEK	1633.82	331	3	-2.20%
T	ALBU_BOVIN	437-451	KVPQVSTPTLVEVSR*	1639.94	349	1	2.99%
T	ALDOA_RABIT	43-56	LQSIGTENTEENRR	1646.81	322	1	-5.50%
P/C	LYSC_CHICK	57-71	NTQATNRNTDGSTDY	1657.70	306	4	-10.44%
T	ALBU_BOVIN	469-482	MPCTEDYLSLILNR	1667.81	361	2	5.27%
C	HBA_HORSE	35-48	GFPTTKTYFPHFDL	1670.82	344	2	0.13%
P	MYG_HORSE	138-151	FRNDIAAKYKELGF	1671.89	351	1	2.04%
T	LYSC_CHICK	116-130	IVSDGNGMNAWAWR	1675.80	344	4	0.03%
C	ALDOA_RABIT	44-57	QSIGTENTEENRRF	1680.79	320	1	-7.12%
C	OVAL_CHICK	268-281	TSSNVMEERKIKVY	1683.87	343	3	-0.57%
T	OVAL_CHICK	127-142	GGLEPINFQTAADQAR*	1687.84	338	2	-2.21%
P	HBA_HORSE	110-125	AVHLPNDFTPAVHASL	1688.88	348	2	0.43%
C	OVAL_CHICK	84-97	RDILNQITKPNDVY	1688.90	341	1	-1.50%
C	ALDOA_RABIT	64-79	TADDRVNPCIGGVILF	1689.86	355	2	2.59%
T	ALDOA_RABIT	243-257	YSHEEIAMATVTALR*	1691.84	400	1	15.48%
P	ALBU_BOVIN	25-38	DTHKSEIAHRFKDL	1696.88	348	2	0.31%
C	OVAL_CHICK	268-281	TSSNVMEERKIKVY MSO: 273	1699.87	339	2	-2.51%
C	HBB_HORSE	131-146	QKVVAGVANALAHKYH	1705.95	372	2	6.82%

Enzyme	Protein	Position	Peptide Sequence	m/z	Ω (Å²)	σ	% Dev.
C	ALDOA_RABIT	1-15	PHSHPALTPEQKKEL	1711.91	361	3	3.37%
C	ALBU_BOVIN	411-424	IKQNCDQFEKLGEY	1714.81	340	3	-2.61%
C	HBB_HORSE	16-31	DKVNEEVGGEALGRL	1714.86	347	3	-0.72%
C	CASB_BOVIN	209-224	QEPVLGPVRGPFPIIV	1718.00	367	2	4.76%
P	CASB_BOVIN	32-46	SSSEESITRINKKIE	1720.91	336	3	-4.06%
C	OVAL_CHICK	199-212	RVTEQESKPVQMMY	1725.83	344	1	-1.96%
C	CYC_HORSE	83-97	AGIKKKTEREDLIAY	1734.97	381	3	8.11%
C	OVAL_CHICK	16-29	KELKVHHANENIFY	1741.90	356	3	0.77%
P	HBA_HORSE	84-98	SDLHAHKLRVDPVNF	1747.92	350	2	-1.19%
C	ALBU_BOVIN	243-257	SQKFPKAEFVEVTKL	1750.97	368	5	3.76%
T	LYSC_CHICK	64-79	NTDGSTDYGILQINSR*	1753.84	342	4	-3.60%
C	HBB_HORSE	13-28	ALWDKVNEEVGGEAL	1758.85	355	3	-0.03%
P	ALBU_BOVIN	553-567	VELLKHKPKATEEQL	1763.01	372	2	4.62%
T	OVAL_CHICK	323-339	ISQAVHAAHAEINEAGR	1773.90	359	2	0.45%
C	CASB_BOVIN	93-108	TQTPVVVPPFLQPEVM	1781.95	360	2	0.48%
P	ALDOA_RABIT	228-243	LKPNMVTPGHACTQKY	1787.89	343	2	-4.41%
T	ENO1_YEAST	358-374	AAQDSFAAGWGVMVSHR*	1789.84	364	3	1.23%
P	LYSC_CHICK	127-141	VAWRNRCKGTDVQAW	1789.89	341	1	-5.16%
T	HBB_HORSE	67-82	VLHSFGEGVHHLDNLK*	1801.93	380	2	5.33%
P	MYG_HORSE	70-86	TALGGILKKKGHHEAEL	1802.03	368	1	2.03%
C	ALDOA_RABIT	63-79	LTADDRVNPCIGGVILF	1802.95	339	2	-6.29%
T	LYSC_CHICK	115-130	KIVSDGNGMNAWVAWR	1803.90	358	3	-1.04%
P	OVAL_CHICK	188-202	FKDEDTQAMPFRVTE	1813.84	346	2	-4.59%
T	MYG_HORSE	1-16	GLSDGEWQQVLNVWGK*	1815.90	363	4	0.11%
T	ENO1_YEAST	375-391	SGETEDTFIADLVVGLR*	1821.92	391	2	7.61%
T	ALBU_BOVIN	508-523	RPCSALTPDETYVPK	1823.90	364	1	-0.01%
C	HBB_HORSE	16-32	DKVNEEVGGEALGRLL	1827.94	364	2	-0.28%
T	HBA_HORSE	41-56	TYFPHFDLSHGSAQVK	1833.89	374	1	2.46%
T	ENO1_YEAST	32-49	SIVPSGASTGVHEALEMR*	1840.92	365	3	-0.43%
C	ALDOA_RABIT	44-58	QSIGTENTEENRRFY	1843.86	342	3	-6.65%
T	ENO1_YEAST	241-257	VKIGLDCASSEFFKDGK	1843.93	364	5	-0.60%
T	ALDOA_RABIT	243-258	YSHEEIAMATVTALRR*	1847.94	370	1	0.80%
C	CYC_HORSE	83-98	AGIKKKTEREDLIAYL	1848.06	387	7	5.38%
T	MYG_HORSE	80-96	GHHEAELKPLAQSHATK*	1853.96	365	4	-0.83%
C	MYG_HORSE	73-89	GGILKKKGHHEAELKPL	1855.09	379	2	2.84%
P	MYG_HORSE	138-153	FRNDIAAKYKELGFQG	1856.97	373	1	1.31%
P	ALBU_BOVIN	431-446	IVRYTRKVPQVSTPTL	1858.09	379	2	2.77%
T	OVAL_CHICK	143-158	ELINSWVESQTNGIIR	1858.97	367	3	-0.33%
C	HBA_HORSE	15-32	SKVGGHAGEYGAEALERM MSO: 32	1861.89	360	3	-2.51%
T	ENO1_YEAST	272-287	WLTGPQLADLYHSLMK*	1872.97	387	2	4.40%
P/C	CASB_BOVIN	208-224	YQEPVLGPVRGPFPIIV	1881.06	381	6	2.53%
T	MYG_HORSE	103-118	YLEFISDAIIHVLHSK*	1885.02	419	4	12.68%

Enzyme	Protein	Position	Peptide Sequence	m/z	Ω (Å^2)	σ	% Dev.
T	ALBU_BOVIN	169-183	HPYFYAPELLYYANK	1888.93	384	1	3.17%
T	CYC_HORSE	89-104	TEREDLIAYKKATNE	1893.99	387	5	3.81%
C	ALBU_BOVIN	377-393	EATLEECCAKDDPHACY	1897.74	382	4	2.13%
P	ALDOA_RABIT	62-79	LLTADDRVNPCIGGVILF	1916.03	362	1	-3.64%
C	MYG_HORSE	12-29	NVWGKVEADIAGHGQEVL	1921.98	374	3	-0.73%
P	LYSC_CHICK	27-43	AAAMKRHGLDNYRGYSL	1922.97	372	1	-1.34%
T	HBB_HORSE	66-82	KVLHSFGEGVHHLDNLK	1930.03	389	3	2.97%
P	ALDOA_RABIT	8-24	TPEQKKELSDIAHRIVA	1935.07	369	2	-2.45%
P	OVAL_CHICK	13-28	DVFKELKVHHANENIF	1940.00	361	2	-4.71%
P	LYSC_CHICK	57-74	NTQATNRNTDGSTDYGIL	1940.89	355	2	-6.27%
P	CYC_HORSE	67-82	YLENPKKYIPGTKMIF	1942.05	385	1	1.56%
C	ALDOA_RABIT	251-269	ATVTALRRTVPPAVTGVTF	1957.12	366	4	-4.01%
C	MYG_HORSE	87-103	KPLAQSHATKHKIPIKY	1960.15	385	1	0.82%
P	LYSC_CHICK	36-52	DNYRGYSLGNWVCAAKF	1963.91	351	2	-8.09%
P	HBB_HORSE	79-96	DNLKGTFAALSELHCDKL	1975.00	382	2	-0.26%
T	MYG_HORSE	79-96	KGHHEAELKPLAQSHATK*	1982.06	386	4	0.46%
C	HBA_HORSE	107-125	STLAVHLPNDFTPAVHASL	1990.04	388	2	0.77%
P/C	CASB_BOVIN	207-224	LYQEPVLGPVRGPFPIIV	1994.15	402	4	4.17%
T	HBB_HORSE	41-59	FFDSFGDLSNPGAVMGNPK*	1999.92	399	1	3.15%
C	OVAL_CHICK	19-35	KVHHANENIFYCPIAIM	1999.99	381	2	-1.28%
C	HBA_HORSE	49-68	SHGSAQVKAHGKKVGDALTL	2004.10	386	2	-0.24%
C	ALBU_BOVIN	487-504	HEKTPVSEKVTKCCTESL	2018.99	374	3	-3.82%
T	ALBU_BOVIN	168-183	RHPYFYAPELLYYANK	2045.03	397	1	1.27%
C	MYG_HORSE	15-33	GKVEADIAGHGQEVLIRLF	2052.12	404	2	2.87%
P	HBA_HORSE	81-98	SNLSDLHAHKLRVDPVNF	2062.08	392	1	-0.50%
T	CYC_HORSE	56-72	GITWKEETLMEYLENPK*	2081.03	416	3	4.99%
C	OVAL_CHICK	126-144	RGGLEPINFQTAADQAREL	2086.07	393	3	-0.88%
C	LYSC_CHICK	82-101	CNDGRTPGSRNLCNIPCSAL	2090.95	381	2	-4.25%
C	LYSC_CHICK	103-123	SSDITASVNCAKKIVSDGNGM	2097.00	386	4	-3.15%
C	ALBU_BOVIN	109-126	GDMADCCEKQEPERNECF	2103.79	380	4	-4.81%
T	ALDOA_RABIT	153-172	IGEHTPSALAIMENANVLAR*	2107.10	419	1	4.78%
P	CASB_BOVIN	206-224	LLYQEPVLGPVRGPFPIIV	2107.23	416	2	4.03%
P	CYC_HORSE	47-64	TYTDANKNKGITWKEETL	2112.06	387	1	-3.20%
C	LYSC_CHICK	53-71	ESNFNTQATNRNTDGSTDY	2134.89	385	3	-4.33%
C	ALBU_BOVIN	557-574	KHKPKATEEQLKTVMENF	2158.13	402	4	-0.93%
C	MYG_HORSE	105-123	EFISDAIIHVLHSKHPGDF	2162.10	420	2	3.52%
T	CASB_BOVIN	199-217	DMPIQAFLLYQEPVLGPVR*	2186.17	433	3	5.85%
C	OVAL_CHICK	41-59	VYLGAKDSTRTQINKVVRF	2195.23	402	3	-1.97%
T	CYC_HORSE	56-73	GITWKEETLMEYLENPKK	2209.12	431	5	4.82%
C	MYG_HORSE	50-69	KTEAEMKASEDLKKHGTVVL	2214.18	423	2	2.72%
P	OVAL_CHICK	151-171	SQTNGIIRNVLQPSSVDSQT A	2215.13	402	2	-2.49%
P/C	HBA_HORSE	47-68	DLSHGSAQVKAHGKKVGDALTL	2232.21	413	3	-0.17%

Enzyme	Protein	Position	Peptide Sequence	m/z	Ω (Å²)	σ	% Dev.
T	ALDOA_RABIT	342-363	YTPSGQAGAAASESLFISNH AY*	2242.04	400	2	-3.71%
C	ALDOA_RABIT	116-137	AGTNGETTTQGLDGLSERCAQY	2272.01	400	2	-4.54%
T	ALDOA_RABIT	111-133	GVVPLAGTNGETTTQGLDGL SER	2272.14	412	2	-1.73%
C	LYSC_CHICK	81-101	WCNDGRTPGSRNLCNIPCSAL	2277.03	407	3	-2.96%
C	CASB_BOVIN	159-178	MHQPHQPLPPTVMFPPQSVL	2281.16	430	1	2.51%
T	OVAL_CHICK	200-218	VTEQESKPVQMMYQIGLFR*	2284.15	431	3	2.62%
C	ALBU_BOVIN	25-43	DTHKSEIAHRFKDLGEEHF	2296.11	413	4	-1.91%
T	HBB_HORSE	9-30	AAVLALWDKVNEEVGGEALGR	2326.20	447	1	5.30%
C	ALBU_BOVIN	308-328	EKSHCIAEVEKDAIPENLPPL	2332.19	415	6	-2.56%
C	CYC_HORSE	11-32	VQKCAQCHTVEKGGKHKTGPN L	2363.21	403	3	-6.04%
C	ALBU_BOVIN	287-307	ICDNQDTISSKLKECCDKPLL	2366.14	414	2	-3.46%
C	ALBU_BOVIN	373-393	AKEYEATLEECCAKDDPHACY	2388.98	421	3	-2.44%
P	ALBU_BOVIN	383-403	CCAKDDPHACYSTVFDKLKH L	2394.10	420	2	-2.94%
C	LYSC_CHICK	53-74	ESNFNTQATNRNTDGSTDYGIL	2418.08	418	3	-3.91%
T	ENO1_YEAST	415-436	IEEELGDNAVFAGENFHHGDKL*	2441.14	434	3	-0.79%
T	OVAL_CHICK	159-181	NVLQPSSVDSQTAMVLVNAIVFK	2460.32	456	5	3.65%
P	LYSC_CHICK	22-43	GRCELAAAMKRHGLDNYRGYSL	2481.22	429	1	-2.87%
P	ALBU_BOVIN	254-275	VTKLVTDLTKVHKECCHGDLLE	2481.28	427	2	-3.27%
C	ALBU_BOVIN	454-475	GKVGTRCCTKPESERMPCTEDY	2490.09	424	4	-4.17%
C	CYC_HORSE	37-59	GRKTGQAPGFTYTDANKNKGIT W	2511.27	418	2	-6.18%
C	HBA_HORSE	43-66	FPHFDLSHGSAQVKAHGKKVGD AL	2546.33	451	3	0.50%
P	HBB_HORSE	63-85	AGKKVLHSFGEGVHHLDNLKGT F	2557.34	450	2	0.03%
C	OVAL_CHICK	149-172	VESQTNGIIRNVLQPSSVDSQTA M	2574.28	447	3	-1.07%
T	MYG_HORSE	97-118	HKIPIKYLEFISDAIIHVLHSK	2601.49	488	2	7.24%
C	ALBU_BOVIN	454-476	GKVGTRCCTKPESERMPCTEDY L	2603.17	438	1	-3.67%
C	CASB_BOVIN	118-140	APKHKEMPFPKYPVEPFTESQSL	2687.35	468	2	1.02%
C	LYSC_CHICK	57-80	NTQATNRNTDGSTDYGILQINSR	2725.29	442	4	-5.44%
C	ALBU_BOVIN	454-478	GKVGTRCCTKPESERMPCTEDY LSL	2803.29	463	4	-2.62%
T	CASB_BOVIN	199-224	DMPIQAFLLYQEPVLGPVRGPFPI IV	2909.60	477	5	-1.74%
T	CASB_BOVIN	199-224	DMPIQAFLLYQEPVLGPVRGPFPI IV MSO: 200	2925.59	472	2	-3.18%
T	HBA_HORSE	62-90	VGDALTLAVGHLDDLPGALSNLS	2949.54	522	2	6.59%
C	CASB_BOVIN	49-73	QSEEQQQTEDELQDKIHPFAQTQ SL	2957.38	482	5	-1.74%
C	OVAL_CHICK	60-87	DKLPGFGDSIEAQCGTSVNVHSS LRDIL	2958.46	485	4	-1.16%
C	CASB_BOVIN	109-134	GVSKVKEAMAPKHKEMPFPKYP VEPF	2971.56	521	5	5.97%
T	ENO1_YEAST	288-311	KYPIVSIEDPFAEDDWEAWSHFF K*	2984.39	488	2	-1.02%
P	CASB_BOVIN	179-205	SLSQSKVLPVPQKAVPYPQRDM PIQAF	3024.63	502	3	1.13%
P	OVAL_CHICK	39-65	AMVYLGAKDSTRTQINKVVRFD KLPGF	3054.66	484	3	-3.13%
P	LYSC_CHICK	75-101	QINSRWWCNDGRTPGSRNLCNIP CSAL	3061.43	482	4	-3.51%

Enzyme	Protein	Position	Peptide Sequence	m/z	Ω (Å²)	σ	% Dev.
C	OVAL_CHICK	145-172	INSWVESQTNGIIRNVLQPSSVDSQTAM	3074.52	493	5	-1.71%
T	HBA_HORSE	61-90	KVGDALTLAVGHLDDLPGALSNI	3077.64	539	3	7.42%
P	CYC_HORSE	37-64	GRKTGQAPGFTYTDANKNKGITWKEETL	3111.59	496	3	-1.76%
T	LYSC_CHICK	52-79	FESNFNTQATNRNTDGSTDYGILQ	3163.47	492	9	-3.33%
P	LYSC_CHICK	75-102	QINSRWWCNDGRTPGSRNLCNIPCSALL	3174.51	496	2	-2.88%
P	CASB_BOVIN	153-180	PLLQSWMHQPHQPLPPTVMFPPQSVLSL	3205.67	526	4	2.39%
P	ALBU_BOVIN	447-475	VEVSRSLGKVGTRCCTKPESERMPCTEDY	3260.52	499	1	-3.65%
C	ALDOA_RABIT	174-203	ASICQQNGIVPIVEPEILPDGDHDLKRCQY	3350.65	526	4	-0.13%
T	ENO1_YEAST	200-233	RYGASAGNVGDEGGVAPNIQTAEEALDLIVDAIK	3413.72	537	4	1.06%
C	CASB_BOVIN	109-140	GVSKVKEAMAPKHKEMPFPKYPVEPFTESQSL	3616.85	572	3	4.19%
T	ENO1_YEAST	141-177	TSPYVLPVPFLNVLNGGSHAGGALALQEFMIAPTGAK	3737.97	584	2	4.41%
P	MYG_HORSE	116-149	HSKHPGDFGADAQGAMTKALELFRNDIAAKYKEL MSO: 131	3745.88	590	2	5.21%
P	HBB_HORSE	46-81	GDLSNPGAVMGNPKVKAHGKKVLHSFGEGVHHLDNL	3759.95	568	2	1.13%
P	HBA_HORSE	1-36	VLSAADKTNVKAAWSKVGGHAGEYGAEALERMFLGF	3780.92	569	1	1.02%
P	ALBU_BOVIN	556-591	LKHKPKATEEQLKTVMENFVAFVDKCCAADDKEACF	4086.98	591	2	0.32%
P	MYG_HORSE	70-106	TALGGILKKKGHHEAELKPLAQSHATKHKIPIKYLEF	4132.35	627	1	5.88%
P	HBB_HORSE	46-85	GDLSNPGAVMGNPKVKAHGKKVLHSFGEGVHHLDNLKGTF	4193.18	608	2	1.85%
P	ALBU_BOVIN	568-607	KTVMENFVAFVDKCCAADDKEACFAVEGPKLVVSTQTALA	4249.07	589	1	-2.24%
P	ALBU_BOVIN	52-90	SQYLQQCPFDEHVKLVNELTEFAKTCVADESHAGCEKSL	4397.05	609	2	-0.78%
P	ALBU_BOVIN	371-410	RLAKEYEATLEECCAKDDPHACYSTVFDKLKHLVDEPQNL	4622.20	641	2	1.32%
P	ALBU_BOVIN	94-136	FGDELCKVASLRETYGDMADCCEKQEPERNECFLSHKDDS PDL	4923.13	647	3	-1.78%
P	LYSC_CHICK	102-147	LSSDITASVNCAKKIVSDGNGMNAWVAWRNRCKGTDVQAW IRGCRL	5050.51	654	3	-2.32%
P	ALBU_BOVIN	308-353	EKSHCIAEVEKDAIPENLPPLTADFAEDKDVCKNYQEAKD AFLGSF	5125.44	668	3	-1.19%

APPENDIX B

THE ION NEUTRAL COLLISION CROSS-SECTIONS OF STANDARD

PEPTIDE IONS MEASURED BY MALDI-IM-TOFMS

Name	Squence	m/z	Average ccs (A^2)	Error ccs (A^2)
Substance P	RPKPQQFFGLM-NH$_2$	1347.7	292	
Gramicidin	HCO-VGALAVVVWLWLWLWLW-NHCH$_2$CH$_2$OH+Na	1905.3	373	
	HCO-VGALAVVVWLWLWLWLW-NHCH$_2$CH$_2$OH	1883.3	368	
BK 1-5	RPPGF	573.3	169	4
BK 2-7	PPGFSP	601.3	174	4
BK 1-6	RPPGFSPFR	660.4	182	5
BK 1-7	RPPGFSP	757.4	197	4
BK 1-8	RPPGFSPF	904.5	223	2
BK 2-9	PPGFSPFR	904.5	215	3
BK 1-9	RPPGFSPFR	1060.6	247	5
Melittin	KRQQ-NH$_2$	558.3	156	5
	GIGAVLK	657.4	189	6
	RKRQQ-NH$_2$	714.4	185	7
	VLTTGLPALISWIK	1511.9	341	4
	VLTTGLPALISWIKR	1668.0	344	10
	VLTTGLPALISWIKRK	1796.1	363	
	VLTTGLPALISWIKRKR	1952.2	383	13
	GIGAVLKVLTTGLPALISWIK	2150.3	401	
	VLTTGLPALISWIKRKRQQ-NH$_2$	2207.3	405	1
	GIGAVLKVLTTGLPALISWIKR	2306.4	422	1
	GIGAVLKVLTTGLPALISWIKRK	2434.5	433	
	GIGAVLKVLTTGLPALISWIKRKR	2590.6	446	10
	GIGAVLKVLTTGLPALISWIKRKRQQ-NH$_2$	2846.5	479	3
ACTH 6-24		2335.9	409	
ACTH 18-39		2465.7	424	
ACTH 1-24		2933.5	461	
ACTH 12-39		3173.9	492	
ATCH 1-17	SYSMEHFRWGKPVGKKR	2093.4	382	1
[Val4] Ang III	RVYVHPF	917.1	226	1
a1-Mating factor 1-6	WHWLQL	882.0	212	9
a-Melanocyte Stimulating Hormone	Ac-SYSMEHFRWGKPV-NH$_2$	1664.9	324	2
Ang I	DRVYIHPFHL	1296.7	288	1
Ang II	DRVYIHPF	1046.5	245	2
Ang II substrate	DRVpYIHPF	1126.2	251	1
Ang III	RVYIHPF	931.5	230	2
Antiflammin-2	HDMNKVLVL	1084.3	244	1
Arg8-Onopressin	CFIRNCPRG-NH$_2$	1062.3	237	3
Flag	DYKDDDK	1013.0	240	3
Interleukin 163-171	VQGEESNDK	1005.0	233	10
Laminin 929-933	YIGSR	594.3	164	1
Neuro 1-24	YPSKPDNPGEDAPAEDMARYYSAL-NH$_2$	2656.8	433	12
Neuro 13-36	PAEDLARYYALRHYINLITRQRY-NH$_2$	2982.4	480	8
Neurotensin	(pGlu)LYENKPRRPYIL	1673.0	333	5
Splenopentin	RKEVY	693.8	176	3
	APGPR	497.3	153	8
	DSDPR	589.3	157	3
	PHCKRM	771.4	199	2

APPENDIX C

T_o VS. M/Z

t_o vs. m/z
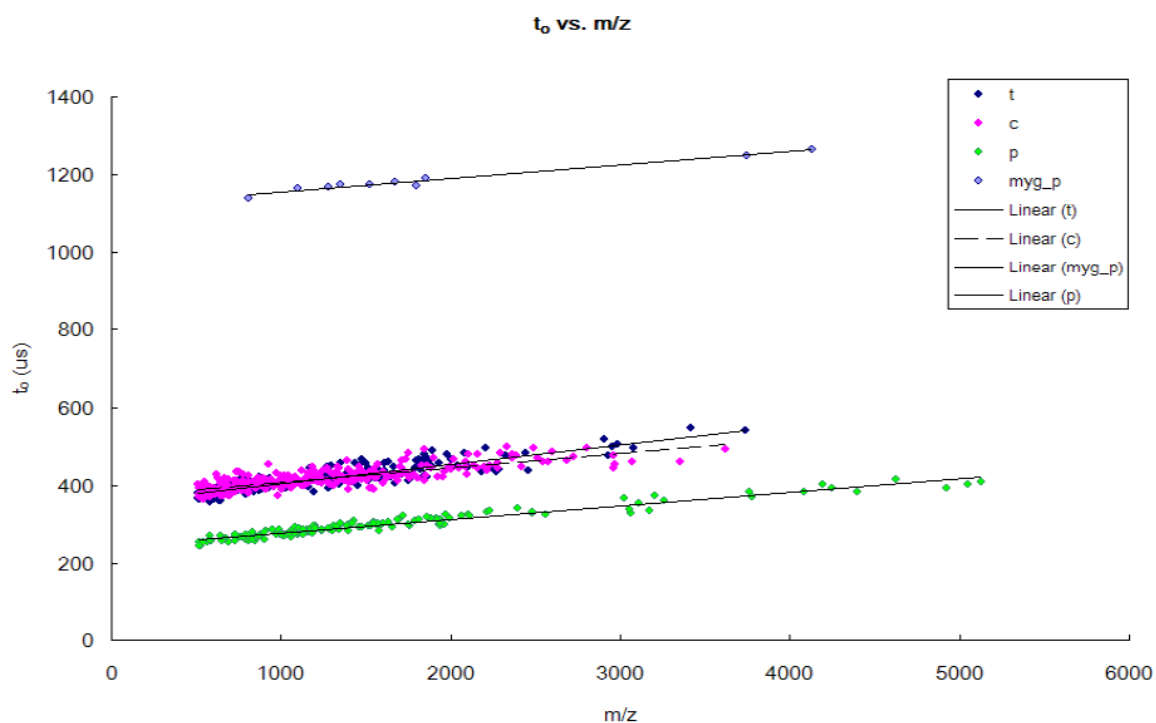
1_1

Tryptic: $t_o = 0.0426(m/z) + 363.37$ $R^2 = 0.7141$

Chymo: $t_o = 0.0426(m/z) + 363.37$ $R^2 = 0.7141$

Peptic: $t_o = 0.0350(m/z) + 242.26$ $R^2 = 0.9406$

Myg_p: $t_o = 0.0344(m/z) + 1121.6$ $R^2 = 0.9738$
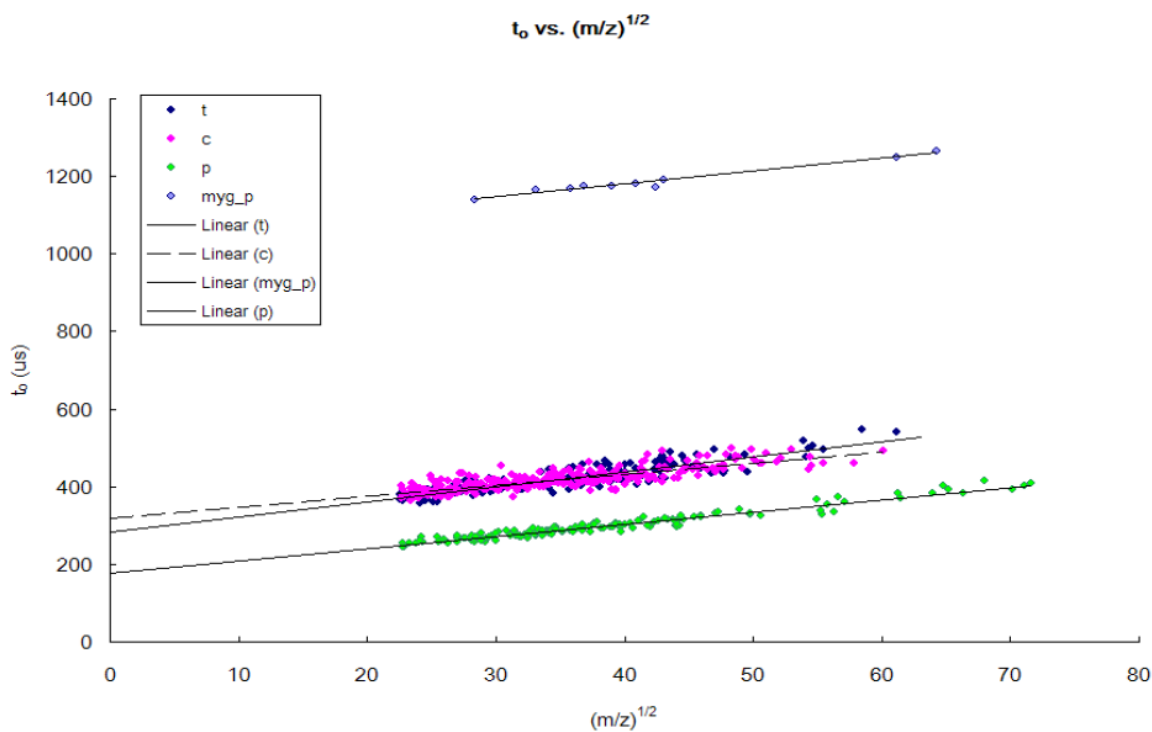
1_2

Tryptic: $t_o = 0.0515(m/z) + 351.02$ $R^2 = 0.7795$

Chymo: $t_o = 0.0381(m/z) + 368.77$ $R^2 = 0.6834$

Peptic: $t_o = 0.0350(m/z) + 242.26$ $R^2 = 0.9406$

Myg_p: $t_o = 0.0344(m/z) + 1121.6$ $R^2 = 0.9738$

APPENDIX D

T_o VS. $(M/Z)^{1/2}$

t_o vs. $(m/z)^{1/2}$

2_1

Tryptic:	$t_o = 3.1996(m/z)^{1/2} + 306.52$	$R^2 = 0.7170$
Chymo:	$t_o = 3.1996(m/z)^{1/2} + 306.52$	$R^2 = 0.7170$
Peptic:	$t_o = 3.1752(m/z)^{1/2} + 176.25$	$R^2 = 0.9468$
Myg_p:	$t_o = 3.3073(m/z)^{1/2} + 1047.4$	$R^2 = 0.9735$

2_2

Tryptic:	$t_o = 3.8556(m/z)^{1/2} + 283.40$	$R^2 = 0.7754$
Chymo:	$t_o = 2.8472(m/z)^{1/2} + 318.38$	$R^2 = 0.6921$
Peptic:	$t_o = 3.1752(m/z)^{1/2} + 176.25$	$R^2 = 0.9468$
Myg_p:	$t_o = 3.3073(m/z)^{1/2} + 1047.4$	$R^2 = 0.9735$

VITA

Name: Lei Tao

Address: Rm 301 Building 2, 778 Wenshui Rd East,
                   Shanghai, China, 200434

Email Address: taolei@gmail.com

Education: B.A., Chemistry, Fudan University, Shanghai, China, 1998
                   M.S., Chemistry, The University of Akron, 2003
                   Ph.D., Chemistry, Texas A&M University, 2010