This is to certify that the

dissertation entitled

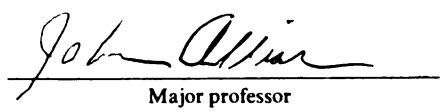
MATRIX-ASSISTED LASER DESORPTION/IONIZATION
MASS SPECTROMETRY:
FUNDAMENTAL STUDIES AND ITS APPLICATION
IN THE ANALYSIS OF PHOSPHOPROTEINS

presented by

Pao-Chi Liao

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Chemistry



Major professor

Date 3-22-95

MATRIX-ASSISTED LASER DESORPTION/IONIZATION MASS SPECTROMETRY: FUNDAMENTAL STUDIES AND ITS APPLICATION IN THE ANALYSIS OF PHOSPHOPROTEINS

By

Pao-Chi Liao

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry

1995

ABSTRACT

MATRIX-ASSISTED LASER DESORPTION/IONIZATION
MASS SPECTROMETRY:
FUNDAMENTAL STUDIES AND
ITS APPLICATION IN THE ANALYSIS OF PHOSPHOPROTEINS

By

Pao-Chi Liao

Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) is a new technique which allows for the analyses of large molecules with the molecular masses up to a few hundred thousand Daltons. While the fundamental aspects of MALDI are still under investigation, it has proven to be a powerful analytical method. This dissertation describes some fundamental studies of MALDI as well as its applications.

The relative signal intensities of $[M+H]^+$ *vs.* $[M+Na]^+$ ions of some small peptides are found to be highly matrix-dependent in MALDI experiments. Presumably, this observation results from the competition of two different ionization mechanisms. Possible mechanisms for the formations of $[M+H]^+$ and $[M+Na]^+$ ions are proposed and discussed. The study suggests that proton transfer from a matrix molecule to an analyte plays an important role in the ionization step. The transferring proton may be derived from photoionized or electronically-excited matrix molecules. In contrast, some data are most consistent with a gas phase mechanism for $[M+Na]^+$ ions.

Most MALDI spectra reported in the literature represent the summation or average of 20-200 mass spectral transients. It is found that the mass spectral transients of MALDI-time-of-flight (TOF)-MS rapidly change with time, and an evaluation of the component spectra may yield different conclusions than evaluation of a single summed spectrum. This tool of time-dependent MALDI is used to investigate details of the power-resolution relationship. The study shows that, by changing how MALDI spectra are constructed, "low power resolution" can be realized in "high power" MALDI experiments.

An approach is developed to locate phosphorylation sites in a phosphoprotein using mass mapping by combining MALDI-TOF-MS with specific enzymatic degradation. To perform mass-mapping, the primary structure of the protein must be known. The mass accuracy requirement of this method is evaluated.

A baculovirus/insect cell expression system, which is used to produce large amounts of the phosphoproteins amenable for phosphorylation site studies, is also described.

To my wife, Yu-Chen

ACKNOWLEDGMENT

I like to thank those who made this dissertation possible. My Mom and Dad are on the top of my thanks list. They have been unconditionally supportive to whatever decisions I made, including that I came to the other side of the earth to pursue an unpredictable adventure which indeed satisfies my curiosity but may not make any sense to them. I met my wife, Yu-Chen, when we entered the Chemistry Department of the Tsing-Hua University. In the junior year, I saw Yu-Chen went to a movie with a decent young man and I suddenly realized that I could not afford to lose the chance to spend my life with her, so I asked for the first date. She was a "good" student and I was not at that time. I started studying hard to ensure we have a future. I do believe that I found the missing part of me and became fully functional and motivated.

It was a great pleasure when I took "Discrete Mathematical Structures" class from Professor Lin-Yu Tseng while he was a PhD student in the Tsing-Hua University. He trained me how to "think" and got me interested in science. Although he does not know any chemistry, I know I would not be here to complete my PhD without his inspiration.

I like to thank to many people in the Mass Spectrometry Facility. I learned the instrumentation of time-of-flight mass spectrometers from Gary Schultz. Mel Micke has been my best source to learn about computers. I will not forget that I visited him in the hospital because he had a false alarm of heart disorder. Mike Davenport taught me electronics and we became good friends. I will not forget his invitation over his house for a "working party" in which we and his brothers built a deck in the backyard of the house. Melinda Berning is the best secretary I ever know and gave me numerous helps. Many

people contributed to an agreeable working atmosphere. Bev Chamberlin's smile and Dr. Z.-H. Huang's hard working are worthy of special notes. I surely thank to the directors of the facility, Professor John Allison, Professor Douglas Gage, and Professor J. Throck Watson, who make this nice environment possible. Doug has been helping me in many ways, I like to give him a special gratefulness.

Among the fellows in Allison group, I give special thanks to Kate Noon and Guy Smith, on the top of my thanks to group members. Kate helped me most in using correct English and shared my excitements and disappoints in my research. Guy is a honest scientist and from whom I learned American history and politics. I also thank my best friends, Chingju Lin, Yu-Ming Kuo, and Han-Hua Yang, who spent numerous wonderful weekends with Yu-Chen and me. Many times I thought I had faced unconquerable difficulties in research, they relieved me from the stress.

For the work described in Chapter 6, I like to thank Professor John Wang, Patty Voss, and Yeuo-Guang Tsay. John opened my eyes to the life science. Patty gifted me another pair of hands which perform various experimental techniques in biochemistry. Many data reported in Chapter 6 were actually done by Patty, she deserves most of the credits of this work. Yeuo-Guang taught me how to play tennis. We had many discussions and I found we have a great deal in common in term of the passions to science.

Finally, I like to thank my mentor Professor John Allison. John has always done his best to help me. As an advisor, he challenged me intellectually and gave me the freedom to pursue what I desired to, even when I was not sure myself where to. As a friend, he gave me good advises and tolerated my stubbornness. I will not forget the time I spent with him for preparing my seminars. It is the time he showed me how to deliver a good presentation as well as his understanding of science. I am proud of being his graduate student and hope that he will be proud of me one day.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BCIP	.....	5-bromo-4-chloro-3-indolyl phosphate p-toluidine salt
bp	.....	base pair
D/I	.....	desorption/ionization
DNA	.....	deoxyribonucleic acid
ESI	.....	electrospray ionization
FBS	.....	fetal bovine serum
FT-MS	.....	Fourier-transform mass spectrometry
FWHM	.....	full width at half maximum
G3	.....	Galectin-3
HEPES	.....	N-[2-hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid]
HPLC	.....	high performance liquid chromatography
I	.....	insulin
IEF	.....	isoelectric focusing
LDI	.....	laser desorption/ionization
MALDI	.....	matrix-assisted laser desorption/ionization
MOI	.....	multiplicity of infection
M_r	.....	relative molecular weight
mRNA	.....	messenger RNA
MS	.....	mass spectrometry
MW	.....	molecular weight
PFU	.....	plaque forming unit
pG-3	.....	phosphorylated G-3
pI	.....	isoelectric point
PMSF	.....	phenylmethylsulfonyl fluoride
p-NBT	.....	p-nitro blue tetrazolium chloride
P_t (m, A)	.....	threshold laser power for analyte, A, and matrix, m.
rG-3	.....	recombinant G-3
RNA	.....	ribonucleic acid
SDS-PAGE	.....	sodium dodecyl sulfate-polyacrylamide gel electrophoresis

Sf21	.....	*Spodoptera frugiperda*
TIC	.....	total ion current
TOF	.....	time-of-flight
TRIS	.....	tris[hydroxymethyl]-aminomethane
UV	.....	ultraviolet
$[M+H]^+$	.....	protonated molecule, M
$[M+Na]^+$	.....	sodiated molecule, M
S/N	.....	signal-to-noise ratio
αCN	.....	α-cyano-4-hydroxycinnamic acid

Chapter 1 - Introduction

I. Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry

During the past two decades, a number of new desorption/ionization (D/I) techniques have been developed to analyze large, nonvolatile, and thermally labile chemical compounds. Among these techniques, electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) show the greatest promise for the mass spectrometric analysis of biopolymers in the mass range between a few thousand and a few hundred thousand Daltons [1].

The first attempts to generate ions of organic molecules by direct laser desorption/ionization (LDI) date back to the early 1970s [2,3]. However, the size of molecules that can be desorbed and ionized was limited to ~ 1,000 Daltons for biopolymers and up to 9,000 Daltons for synthetic polymers. When the size of the molecules increases, more energy is required to be deposited into the sample to desorb them. However, such a high energy flux leads to pyrolytic or photochemical decompositions of analytes. The main breakthrough toward higher masses came in 1987 when Hillenkamp and Karas successfully experimented with the use of a matrix, nicotinic acid (3-pyridine carboxylic acid) [4,5]. In their initial experiments, the sample molecules were mixed in solution with a large molar excess of UV-absorbing nicotinic acid matrix and the mixture was deposited and air dried on a metal substrate. The deposit was then desorbed/ionized by a pulsed, frequency-quadrupled 266-nm Nd:YAG laser. The ejected ions were analyzed by a time-of-flight (TOF) mass spectrometer. They showed that the ions generated from

proteins with molecular mass range 10,000-67,000 Daltons could be formed and detected from picomolar amounts of analyte.

After the pioneering discovery by Hillenkamp and Karas, a number of research groups joined this research area. Among them, R. Beavis and B. Chait made important contributions by the introduction of new matrices [6,7] and expanding the range of usable laser wavelengths [8]. They discovered two new matrices, sinapinic acid and α-cyano-4-hydroxycinnamic acid, which can be coupled to the use of 337-nm radiation from a relatively inexpensive nitrogen laser. Today, the most commonly used matrices for analyses of peptides and proteins are sinapinic acid, α-cyano-4-hydroxycinnamic acid, and 2,5-dihydroxybenzoic acid [9]. Although peptides and proteins are the main biocompounds investigated by MALDI so far, the technique also can be applied to oligonucleotides [10-14], oligosaccharides [15-17], glycoconjugates [18,19] and synthetic polymers [20-22].

Figure 1.1 shows a schematic drawing of the VT2000 (Vestec Corp., Houston, TX) linear MALDI-TOF mass spectrometer that has been used to perform many of the experiments described in this dissertation. Due to the pulsed nature of the laser ionization source, TOF appears to be the best choice of mass analyzer for MALDI. The unique, theoretically unlimited, mass range of TOF gives the capacity to analyze large ions with masses higher than 10,000 Daltons. This parallels one of the main benefits of MALDI, the capacity to generate ions from large biomolecules. Although recent advances in Fourier-transform mass spectrometry (FT-MS) have shown great potential in analyzing high-mass ions (i.e., >10,000 Daltons), its practical use is still currently under rapid development [23-25]. Another advantage of the TOF mass analyzer is its capacity to generate the entire spectrum for every single laser shot without losing information, in contrast to scanning mass analyzers. The initial kinetic energy spread of ions generated by MALDI is found to be large [26], so either a linear TOF with a high accelerating voltage

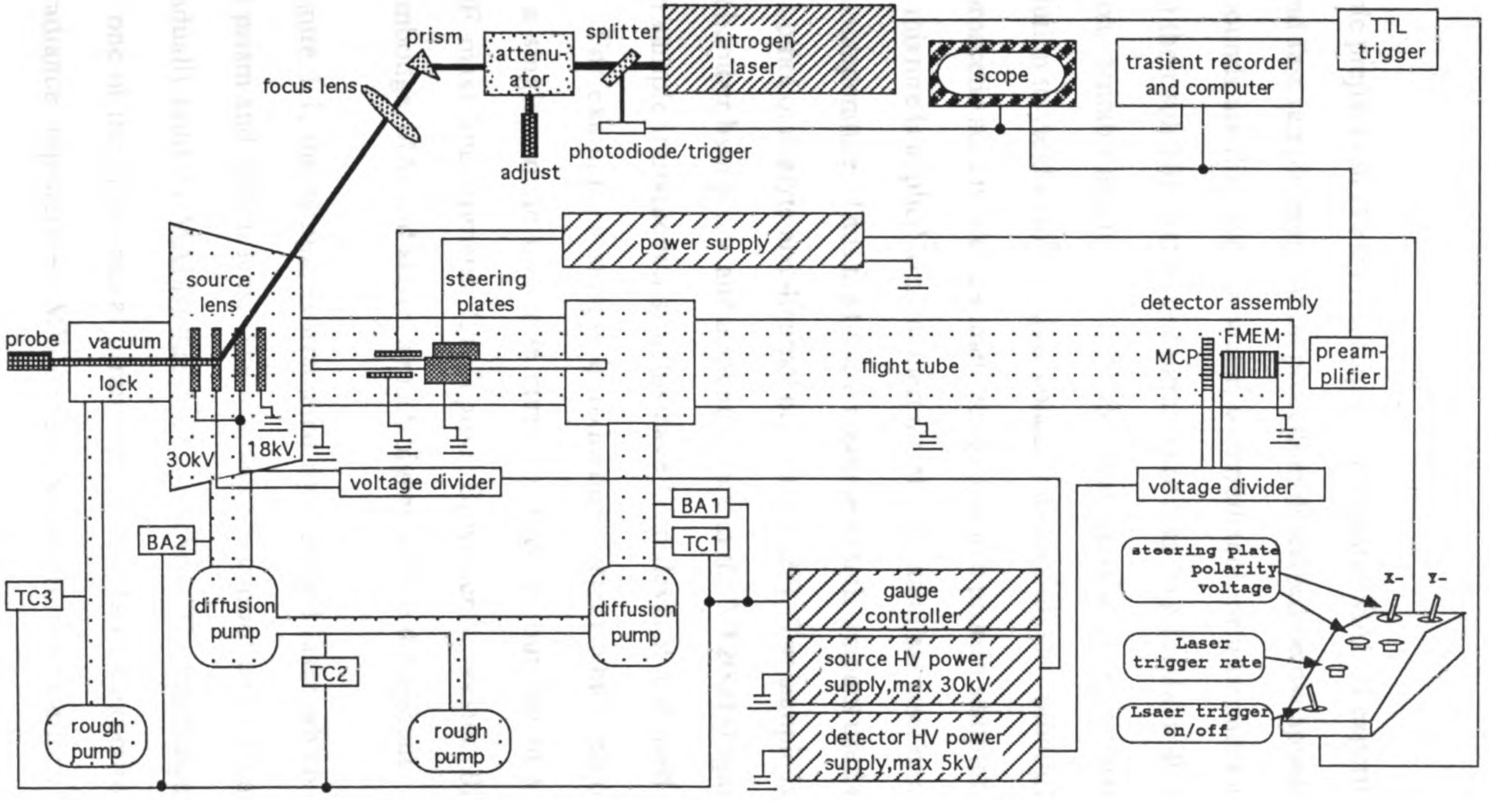


Figure 1.1 A schematic drawing of the VT 2000 linear MALDI-TOF mass spectrometer. Abbreviations used: TC, thermocouple pressure gauge; BA, Bayard-Alpert type pressure gauge; MCP, microchannel plate; FMEM, focused-mesh electron multiplier; TTL, transistor-transistor logic.

(~30 kV) or a reflectron with an ion mirror (or both) is used to improve mass resolution. The VT2000 is a linear MALDI-TOF mass spectrometer equipped with a 30-kV source (Figure 1.1).

Sample preparation is a critical step in a successful MALDI experiment. However, the simple and fast sample preparation procedure described below appears to work well. Matrix compounds are dissolved in water, water/ethanol, or water/acetonitrile mixtures, depending on their solubility, at a concentration of 5 to 10 g/L (about 40 mM) to give the matrix solution. Suitable amounts of 1-10 μM solutions of analyte are mixed with 1-10 μL of matrix solution to yield a final analyte concentration of about 1 pmol/μL in the mixture (0.1% trifluoroacetic acid is added sometimes to dissolve protein samples). An aliquot of ~1 μL of the mixture is applied to a flat inert metal (e.g., stainless steel or silver) probe and dried at room temperature. Thus the typical absolute sample amount used for analysis is in the range of 1 pmol of analyte and 40 nmol of matrix [1,27]. The sample is introduced into the vacuum chamber by a probe and a vacuum lock in the VT2000 (Figure 1.1). Several sophisticated sample introduction systems have been devised and used by commercial instruments. For example, a sample introduction system which accommodates 100 samples on a single sample holder is currently used in our lab in a Voyager Elite MALDI-TOF mass spectrometer (PerSeptive Biosystems, Vestec BioSpectrometry Products, Cambridge, MA) and allows a much higher sample throughput.

In Figure 1.1, the sample is irradiated by a pulsed laser which is directed and focused by a prism and optical lens. The irradiance is controlled by an attenuator and is increased gradually until the MALDI threshold is reached. The irradiance of the laser has proven to be one of the most critical parameters in the MALDI experiment. There is a threshold irradiance, typically ~1 MW/cm^2, necessary to produce ions from a sample. Below this threshold, ion production falls off to the fifth power of laser irradiance [26]. Laser irradiances higher than the threshold dramatically decrease the mass resolution in the

spectra obtained. The size of the laser spot and the angle of incidence of the laser beam on the sample surface seem to not be very critical. Laser spot sizes of 10-300 μm diameters and 30°-75° angles of incidence are typical. For focusing of the laser beam, a single quartz lens of long focal length (10-30 cm) may be used [1].

P. Demirev and his coworkers reported the desorption/ionization of proteins from different matrices by using an excimer-pumped dye laser generating ultra-short (560 fs full width at half maximum, FWHM) pulses in the ultraviolet (UV, 248 nm) and visible (496 nm) spectral ranges [28]. The existence of a threshold energy density for the MALDI of insulin by ultra-short pulses has been established. These data are compared to the data obtained by employing a nitrogen laser (337 nm, 3 ns FWHM). The threshold energies from the two laser systems employed are of the same order of magnitude and do not depend on the laser pulse length (the energy deposition time). Thus P. Demirev *et al.* have concluded that the amount of laser energy deposited into the sample, rather than the laser power, is the important parameter in the MALDI process. The VT2000 employs a nitrogen laser generating 3-ns pulse of radiation at 337 nm (Figure 1.1).

In Figure 1.1, the ions generated from MALDI are accelerated by a two-stage 30-kV electric field toward a flight tube (field-free region). Because all ions have the same kinetic energy gained from the accelerating electric field (assuming that initial kinetic energy is zero), their velocities vary inversely with their masses, with lighter ions traveling faster than heavier ones. Ions with different masses have different flight times, t, which are related to the masses, m, by the following equation:

$$t = (\text{time in ion source}) + (\text{time in flight tube}) = \sqrt{\frac{2dm}{zeV}} + L\sqrt{\frac{m}{2zeV}}$$

where V = accelerating voltage, V/d = electric field in ion source, z = the number of charges, e = 1.609 x 10^{-19} C, and L = the length of the flight tube. In practice, the mass of an ion species is not calculated by the above equation. Instead, it is achieved by

calibration. The square root of m/z is proportional to the flight time for a given mass spectrometer with fixed values of d, V, and L. That is,

$$\sqrt{\frac{m}{z}} \propto t ,$$

$$\text{or, } \sqrt{\frac{m}{z}} = c_1 \times t$$

The flight time, t, is the duration between when the ions were generated and when they reach the detector. However, the actual starting time is difficult to measure, so a reference point in time is registered by a photodiode which receives a split laser beam (Figure 1.1). The time relative to this reference point, t', and the actual flight time, t, differ by a constant value, c_2. Therefore,

$$\sqrt{\frac{m}{z}} = c_1 \times t' + c_2$$

At least two known points of ($\sqrt{\frac{m}{z}}$, t'), usually obtained by the use of standard samples, are necessary to calculate the values of c_1 and c_2.

The most common type of detector is composed of a conversion electrode and an electron multiplier. The former generates electrons from impinging ions. The latter multiplies the number of electrons to generate a larger electrical current. This configuration of detector is also used by the VT2000, where the conversion electrode is a microchannel plate and the electron multiplier is a focused-mesh electron multiplier (Figure 1.1). The main problem in ion detection is that the high-mass ions generated by MALDI move too slowly to generate secondary electrons efficiently at the conversion electrode. The yield of secondary electrons from the conversion electrode is a function of the momentum of the ions detected [1,29]. A high acceleration voltage in the ion source or post-acceleration before the detector may be used to increase the momentum of ions and thus increase the

detection efficiency. Post-acceleration is often used in reflectron-TOF instruments, with typical accelerating voltages of 3~5 kV [30,31].

MALDI spectra obtained by a linear TOF mass spectrometer are dominated by peaks representing the intact protonated molecules. Peaks corresponding to fragmentation of these large ions due to cleavage of covalent bonds in the protein backbone are rarely observed in the spectra. Fragment ions due to the loss of small neutral molecules (such as H_2O, NH_3 and HCOOH) from protonated peptides and proteins are usually of low abundance. The lack of fragmentation and the dominant singly-charged protonated molecule peaks in MALDI spectra make this technique ideal for mixture analysis [27,31,32]. Later studies using a reflectron TOF mass spectrometer reveal that MALDI ions do undergo postsource fragmentations [33,34], but they are not separated from the protonated molecules by a linear TOF mass spectrometer. Peaks representing protonated molecules are accompanied by satellite peaks at higher masses, which can be attributed to addition of matrix molecules or, more often, parts of the matrix molecules. For example, signals are observed for $[M+207]^+$ for sinapinic acid (MW = 224) and $[M+136]^+$ for 2,5-dihydroxybenzoic acid (MW = 153) [27]. The accessible mass range for MALDI has been extended to ~300,000 Daltons [35,36]. In the high-mass range, the satellite peaks due to adduct ions between analyte and matrix cannot be well resolved from the quasimolecular ion peak. This will degrade the mass determination accuracy to some extent. Considerable improvements have been achieved by the discovery of new matrices which yield fewer adduct ions.

MALDI is a very sensitive technique. Typically no more than 1 pmol of analyte is necessary to give a satisfactory MALDI spectrum [37]. If the analyte concentration is too high, a decrease in signal quality is often observed. R. Beavis and B. Chait found that, by using the sinapinic acid as matrix, even a large excess of inorganic salts, denaturation agents (like urea and guanidinium hydrochloride, up to 2M in the protein solution),

commonly used buffering agents (citrate, glycine, N-[2-hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid] (HEPES), tris[hydroxymethyl]-aminomethane (TRIS), ammonium bicarbonate and ammonium acetate up to 200 mM) do not interfere with the formation of protein ions [8,27,32]. Thus elaborate sample purification is not necessary.

II. The Focus and Organization of the Dissertation

The introduction of MALDI has opened a new field for the application of mass spectrometry to the problems of biological interests. MALDI offers certain advantages, such as high mass range, high sensitivity, tolerance to impurities and mixture analysis capability, some of which were not even possible to achieve before the discovery of MALDI. MALDI has become a very useful technique, however, its ionization mechanism is not fully understood and is still a matter of considerable debate and research.

My dissertation research focused on MALDI-TOF-MS. The research interest has been twofold: fundamental studies and applications. I have done some work and tried to understand the chemistry which leads to the ionization process, described in Chapters 2 and 3. I have also discovered that dissecting MALDI mass spectra reveals some interesting information. The relationship between laser power and resolution has been investigated using the tool of dissecting MALDI mass spectra, which is described in Chapter 4. The ultimate goal of these studies is to improve the technique by defining and understanding its chemistry.

Although intense research activities are still ongoing to study the fundamental aspects of MALDI, it has offered many research opportunities - application of this new technique to problems that previously were not possible to solve by mass spectrometry. I have developed a method to locate phosphorylation sites in a phosphoprotein using mass-mapping through combining specific enzymatic degradation with MALDI-MS, which is described in Chapter 5. Through an arrangement made by Professor Douglas A. Gage,

Professor John L. Wang in the Biochemistry Department became interested in the method that I have developed. Professor Wang wanted to apply my method to study a phosphoprotein which he had been working on, but he needed a graduate student to prepare phosphoprotein samples using molecular biology methods. I was interested in the proposed project and also was encouraged by Professor John Allison. So I went to Professor Wang's lab and used a baculovirus/insect cell system to successfully express the target protein. This experience was very valuable to me. It enabled me to communicate effectively with biochemists and broadened my research interests into the life sciences. Although the project has not been completed, this stimulating adventure is discussed in Chapter 6.

Since a part of the dissertation research has been published, the publication reprints are included. The information already contained in the reprints will not be restated in the chapters. The reprints will be mentioned in the proper chapters and will appear as Appendices.

Chapter 2 - Ionization Mechanisms in Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry

Several possible ionization mechanisms for the MALDI process, including those proposed in the literature, have been postulated and evaluated. The work has been submitted and accepted for the publication. The preprint is included as Appendix I. The relative signal intensities of $[M+H]^+$ *vs.* $[M+Na]^+$ ions of some small peptides are found to be highly matrix-dependent in matrix-assisted laser desorption/ionization (MALDI) experiments when sinapinic acid, α-cyano-4-hydroxycinnamic acid and 2,5-dihydroxybenzoic acid are used as matrices. Presumably, this observation results from the competition of two different mechanisms for the ionization steps. Both the formations of protonated and sodiated molecules were discussed in Appendix I. The results suggest that proton transfer from a matrix molecule to an analyte plays an important role in the ionization step. The transferring proton may be derived from photoionized or electronically-excited matrix molecules. In contrast, some data are most consistent with a gas phase mechanism for $[M+Na]^+$ ions.

I have also discovered enhanced detection of peptides in MALDI-MS through the use of charge-localized derivatives. Triphenylphosphonium derivative is one of them. Some peptides do not yield signals in MALDI experiments, but their triphenylphosphonium derivatives do. This finding was initially submitted as a part of the publication in Appendix I. The reviewers suggested that it should be published separately. A preprint of the resulting publication is included as Appendix II.

In addition to its potential use as an analytical method, the precharged triphenylphosphonium derivatives may be used as a probe to study independently the factors contributing to the desorption process and those contributing to the ionization process. In Chapter 3, a survey of new MALDI matrices is described, aiming to understand how and why some compounds function as successful matrices. While trying to deduce some rules accounting for desorption and/or ionization processes, I realized a fundamental difficulty in my general approach: both desorption and ionization must occur to observe signals in MALDI. The precharged triphenylphosphonium derivatives, on the contrary, only require desorption. A compound that is not a working MALDI matrix for a peptide may work for its triphenylphosphonium derivatives because only desorption is required. Such compounds may allow us to deduce some rules accounting for the desorption process alone.

Chapter 3 - A Survey for New MALDI Matrices

I. Introduction

The search for new MALDI matrices has been an important research topic since the discovery of the technique. Only a few compounds are useful as matrices for protein and peptide analyses. R. C. Beavis pointed this out in a review article in 1992: "Out of more than 300 materials tested (to my knowledge), only seven matrices that are practically useful have been found." Up to date, there have not been many matrices reported in the literature. The most commonly used three matrices are sinapinic acid, 2,5-dihydroxybenzoic acid and α-cyano-4-hydroxycinnamic acid, which were discovered in 1989 [6], 1991 [9] and 1992 [7], respectively. Adding new chemical compounds to the list of successful MALDI matrices may not only improve the practical aspects of the technique, but form a basis to give researchers hints on how and why these compounds function as successful matrices. The search for new MALDI matrices taken here was mainly motivated by the latter issue. In consideration of this, the scope of chemical compounds being tested is relatively broad and extensive. The procedures for testing the physical properties of matrix compounds, representative mass spectral data by using successful MALDI matrices, and their implication are reported here.

II. Experimental

Insulin was used as an analyte to test whether a compound is a successful MALDI matrix. It was dissolved in water/acetonitrile (2:1) to give a concentration of 100 pmol/μL. The test for successful matrices was herein limited to peptides. Compounds to be tested as matrices were dissolved in water/acetonitrile (2:1 or 1:1) or acetone. About 30 μg of matrix

in 1-μL volume (roughly 100 nmol, depending on the molecular weight, and if the solubility is too small to completely dissolve the matrix using any of three solvent systems, then saturated solutions of highest concentration using one of three solvent systems was used) was mixed with an equal volume containing 100 pmol of insulin, applied to a stainless probe, dried and analyzed by MALDI-TOF-MS. Analyte and matrix must co-dissolve in the solvent system. The instrument and experimental conditions were the same as described in Appendix I except that the laser power may be much higher when it is necessary to meet the threshold condition for certain matrices. The threshold laser power will be abbreviated as P_t (m, A) for later discussions, where P_t is the threshold power at which the analyte, A, generates observable signals (S/N > 5) while power increases using the matrix, m. Therefore, 10 times of P_t (m, A) will be abbreviated as 10 P_t (m, A). The appearance of signal representing "insulin ions", was used to indicate a successful MALDI matrix. The generic term, "insulin ions", is used because the MALDI-TOF-MS used for these experiments does not provide enough mass resolution and mass accuracy to differentiate between protonated insulin molecules, insulin radical cations, and sodiated insulin molecules.

Molar absorptivities at 337 nm were measured for the tested compounds in dilute methanol solution using a Hitachi UV spectrophotometer (model U-2000). The concentrations of solutions varied from 0.02 to 2 mM.

III. Results and Discussion

A. Chemical compounds which have been tested

The chemical compounds tested were selected according to the criteria described below from the shelves in the laboratory rather than purchased or synthesized. They have to be solid, or at least with low vapor pressure, to sustain in vacuum and codissolve with insulin in a solvent system to validate the test. Most of the compounds selected for testing

were organic compounds and UV chromophores. The chemical compounds which have been tested are tabulated in Table 3.1 with their physical properties, such as molecular weight, melting point and molar absorptivity, as well as the results of the test. If a compound is reported to sublime or decompose at its melting point under atmospheric pressure or in vacuum, it is specified. Accurate UV absorption coefficients for solids are difficult to measure. The molar absorptivities of compounds were measured in dilute methanol solution at 337 nm. For those compounds which function as successful MALDI matrices, the mass spectral peaks representing insulin ions were reported in terms of peak intensity and peak resolution. The peak intensities were classified as strong, medium, weak and very weak. The resolution was defined by $m/\Delta m$ (FWHM).

It is generally believed that successful MALDI matrices are strong absorbers for the laser radiation used in the MALDI experiment [1,27,38]. The wavelength of the nitrogen laser used in this experiment was 337 nm. In Table 3.1, a comparison between the molar absorptivities of successful MALDI matrices and those of non-successful matrices roughly agrees with this trend, and most successful MALDI matrices show high molar absorptivities at 337 nm. However, in some cases, the successful MALDI matrices have low molar absorptivities. For examples, ninhydrin and salicylic acid have low molar absorptivities of 38 and 50 $L \cdot cm^{-1} mole^{-1}$, respectively, but both matrices gave very strong peaks representing insulin ions, indicating that they are successful MALDI matrices. When ninhydrin was used as matrix, very intense laser power, about 10 x P_t (sinapinic acid, insulin), was necessary to generate insulin ions. On the other hand, a strong absorbance alone cannot make a successful MALDI matrix. For example, the molar absorptivity of 1,4-diphenyl-1,3-butadiene was measured to be as high as 27000 $L \cdot cm^{-1} mole^{-1}$, but it is not a successful MALDI matrix. It was also found that most of the successful MALDI matrices sublime or decompose at their melting points, which may indicate they possess the characteristics that facilitate the physical desorption process.

Table 3.1 Chemical Compounds tested as MALDI matrices

Matrix	M.W.	m.p.(C)	molar abs.£§	insulin signal intensity	insulin signal Resol'n
2,5-Dihydroxybenzoic Acid	154.1	200 (sub.)$	2500	****¥	330
alpha-Cyano-4-hydroxycinnamic acid	189.2	251 (dec.)$	1600	****¥	360
Anthranillic acid	137.1	145	4600	****¥	410
Sinapinic acid	224.2	204 (dec.)	13000	****¥	520
2,4-Dihydroxybenzophenone	214.2	144	5000	****	280/80
5-Methylsalicylic acid	152.2	152	510	****¶	240
Ninhydrin	178.1	241 (dec.)	38	****¶	205
p-Nitrophenol	139.1	113 (sub.)	4400	****¶	60
Salicylic acid	138.1	76 (sub.dec.)	50	****¶	200
9-Anthracenecarboxylic acid	222.2	220 (dec.)	-	***	100
N-Phenylmaleimide	173.2	89	< 50	***	90
2-Naphthol	144.2	122 (sub.)	410	***¶	310/175
4-Hydroxybenzaldehyde	122.1	116 (sub.)	200	***¶	150
m-Hydroxybenzoic acid	138.1	202	< 50	**	150
Naphthacene	228.3	341 (sub. in vacuum)	1400	**	50
p-Toluenesulfonic acid	172.2	106	< 50	**	500
3-Aminopyrazine-2-carboxylic acid	139.1	205 (dec.)	4100	**¶	290
1,2-Naphthoquinone	158.2	146 (dec.)	1400	*	63
1-Chloroanthraquinone	242.7	159	2600	*	5
2-methylanthraquinone	222.2	177 (sub.)	4100	*	20
3,5-Dinitrobenzoic acid	212.1	207 (sub.)	130	*	10
Anthracene	178.2	218 (sub.)	2100	*	400/240
Boric acid	61.8	171	< 50	*	180
Caffeine	194.2	178 (sub.)	< 50	*	75
Carbazole	167.2	245 (sub.)	1400	*	35

Table 3.1 Chemical Compounds tested as MALDI matrices (continued)

Matrix	M.W.	m.p.(C)	molar abs.£§	insulin signal intensity	insulin signal Resol'n
DL-Valine	117.2	315 (sub.)	< 50	*	35
Naphthalene	128.2	80.2 (sub.)	< 50	*	35
Phenanthraquinone	208.2	206	1700	*	5
1,2,3-Trimethoxybenzene	168.2	44	< 50	X	-
1-Naphthaleneacetic acid	186.2	130	-	X	-
2'-Cytidylic acid	323.2	239 (dec.)	420	X	-
2,4-Dihydroxybenzoic acid	154.1	226 (dec.)	< 50	X	-
2-Chloro-5-nitropyridine	158.5	108	140	X	-
2-Methylanthraquinone	224.2	171(sub.)	-	X	-
3,4,5-Trimethoxybenzoic acid	212.2	169	< 50	X	-
4-Phenylurazole	177.2	208	< 50	X	-
Acetylsalicylic acid (Aspirin)	180.2	139	< 50	X	-
Anthraquinone	208.2	380(sub. in vacuum)	2900	X	-
Aurintricarboxylic acid, ammonoum salt	473.4	222 (dec.)	-	X	-
Benzil	210.2	94	110	X	-
Benzoic acid	122.1	122(sub.)	< 50	X	-
Benzophenone	182.2	50	71	X	-
Brilliant green	482.6	-	1200	X	-
Bromothymol blue(Na salt)	646.4	-	4400	X	-
d-Mandelic acid	152.2	121	< 50	X	-
Dihydroxyphenylalanine (DOPA)	197.2	290 (dec.)	< 50	X	-
DPB,1,4-Diphenyl-1,3-butadiene	206.3	152	27000	X	-
Histamine dihydrochloride	184.1	250	610	X	-
Hydroquinone	110.1	173	< 50	X	-
KMnO4	158.0	-	< 50	X	-

Table 3.1 Chemical Compounds tested as MALDI matrices (continued)

Matrix	M.W.	m.p.(C)	molar abs.£§	insulin signal intensity	insulin signal Resol'n
L-Tryptophan	204.2	282 (dec.)	1600	X	-
m-Chlorobenzoic acid	156.6	156	-	X	-
Phenol	94.1	41	< 50	X	-
Phthalic anhydride	148.1	132(sub.)	< 50	X	-
Picric acid (explosive!)	229.1	122	6100	X	-
Potassium hydrogen phthalate	204.2	297	< 50	X	-
Pyrene	202.3	150(sub.)	4100	X	-
Quinone	108.1	114(sub.)	160	X	-
Sulfanilic acid, 4-NH2C6H4SO3H	173.2	325(dec.)	19	X	-
Thymine	126.1	316(dec.)	< 50	X	-
Violuric acid, monohydrate	175.1	247 (dec.)	260	X	-

****: strong insulin signal
***: medium insulin signal
**: weak insulin signal
*: very weak insulin signal
X: no insulin signal

£: unit, L/(cm x mole)
§: the concentrations of solutions vary from 0.02 to 2 mM, not shown here
$: abbreviations: sub., sublimation; dec., decomposition.
¥: already reported in literature
¶: strong matrix adduct

B. Examples of new successful MALDI matrices

Several compounds have been newly discovered to be successful MALDI matrices which were not reported in the literature.

1. *2,4-Dihydroxybenzophenone*

Figure 3.1 shows the MALDI spectrum resulting from the mixture of insulin and 2,4-dihydroxybenzophenone. The peak representing the insulin ion is very intense but has a very poor resolution (~ 80) using this high laser power. When the laser power was lowered, the resolution was improved (~ 280) with a less intense peak. No significant matrix adduct ions were observed. A matrix adduct ion is formed by the addition of matrix molecules or parts of the matrix molecules to an insulin ion.

2. *5-Methylsalicylic acid, ninhydrin, p-nitrophenol, and salicylic acid*

These four compounds were successful MALDI matrices and produced intense peaks representing insulin ions, but also generated multiple strong matrix adduct peaks at the higher mass end of the spectrum. These adduct ions may be due to the photochemical reactions of insulin and matrix molecules. The mass differences between insulin ions and adduct ions may not necessarily match multiples of the mass of the matrix molecules, indicating the reactions may be more than simple additions of matrix molecules to insulin ions. Figure 3.2(a) shows the MALDI spectrum resulting from the mixture of insulin and 5-methylsalicylic acid. The laser power used to obtain Figure 3.2(a) was relatively low, comparable to the P_t (sinapinic acid, insulin), to yield better resolution with a less intense peak. When higher laser power was used to obtain Figure 3.2(b), the peaks representing insulin ions became more intense with lower resolution.

Among these four successful MALDI matrices, ninhydrin was a special case because a "preheat process" by laser radiation was necessary prior to the MALDI process

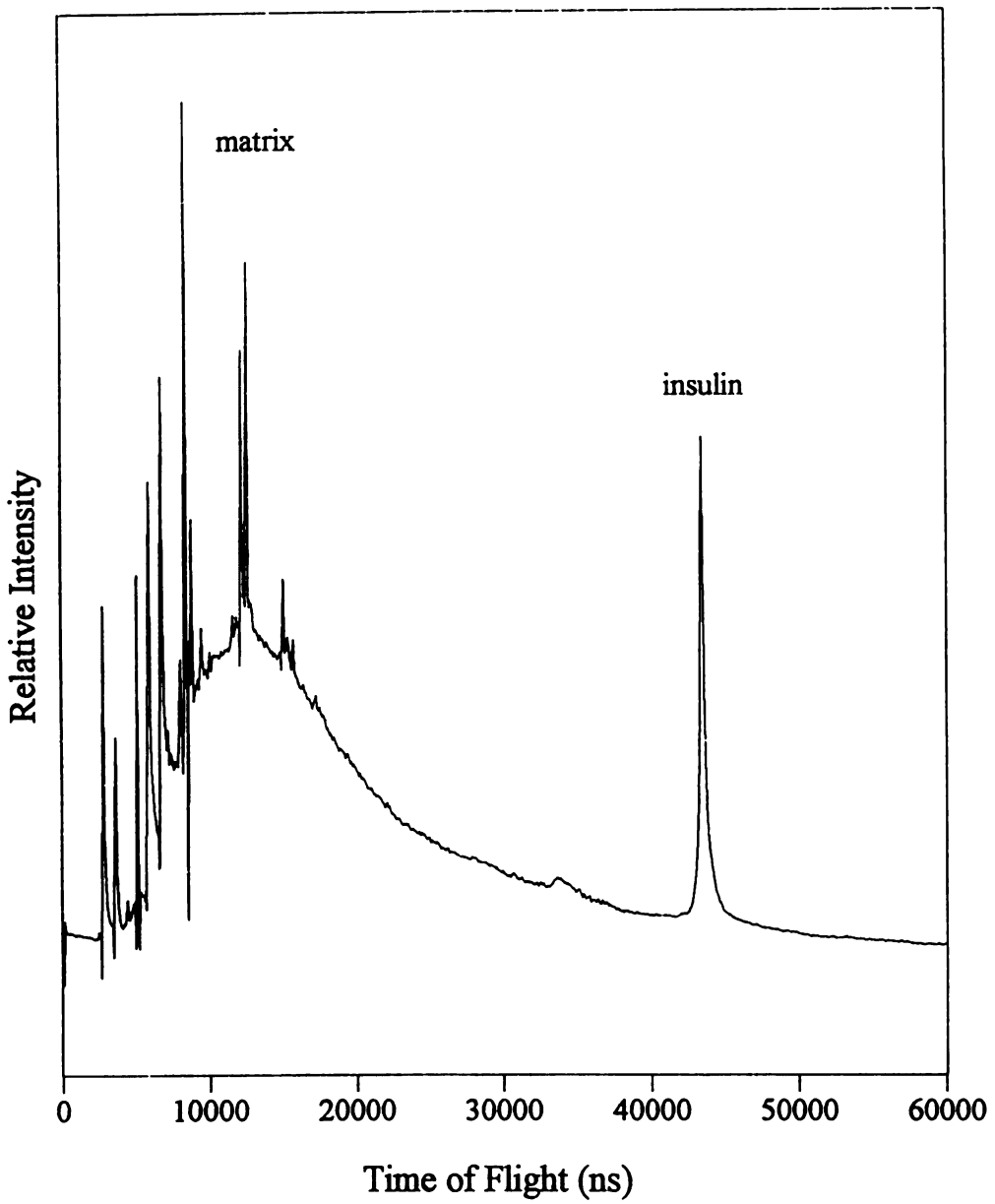


Figure 3.1 The MALDI spectrum resulting from the standard mixture of insulin and 2,4-dihydroxybenzophenone matrix (molar ratio ~ 1: 1000).

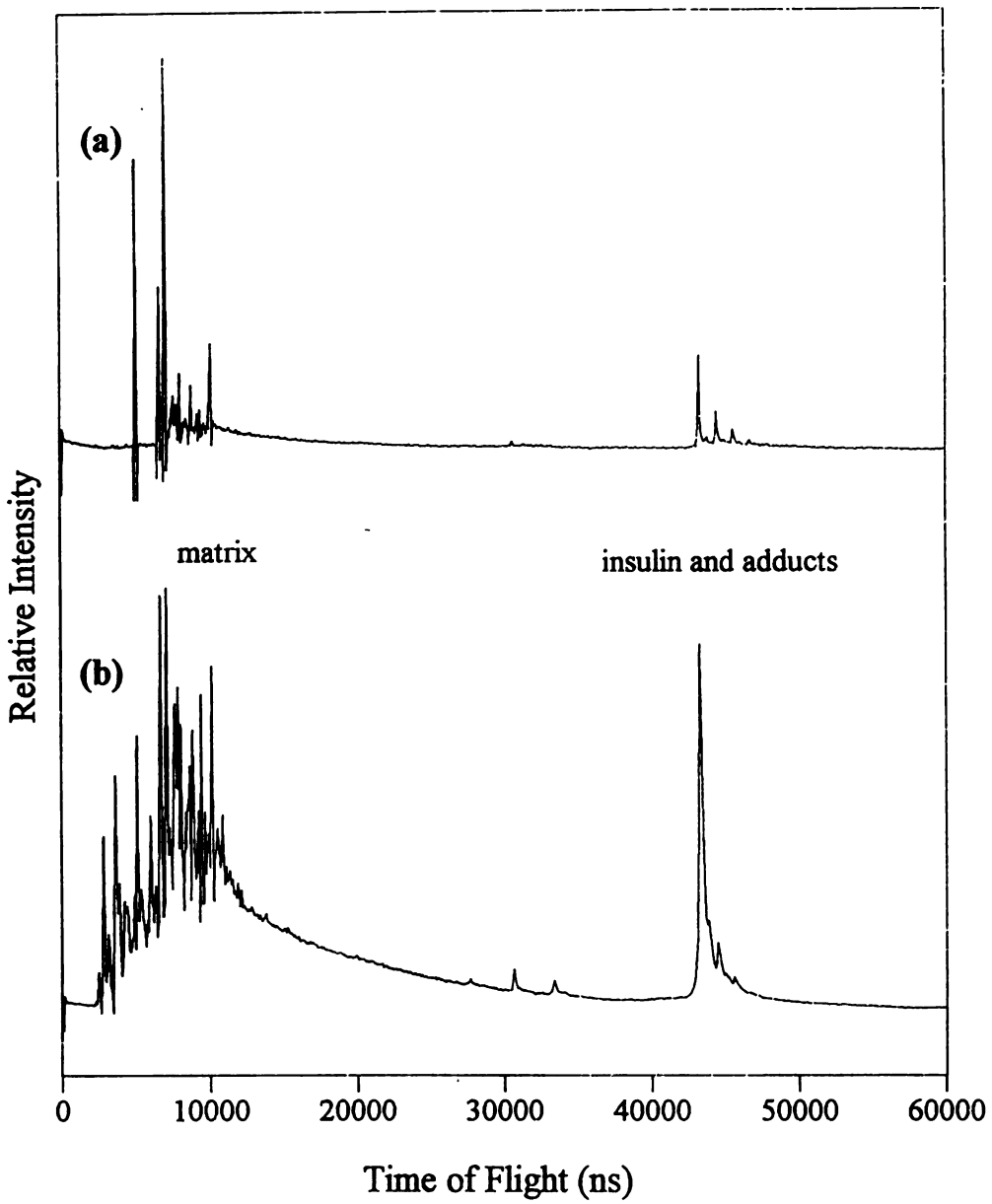


Figure 3.2 The MALDI spectra resulting from the standard mixture of insulin and 5-methylsalicylic acid matrix (molar ratio ~ 1: 1000) using a laser power (a) comparable to, and (b) higher than the Pt (sinapinic acid, insulin).

for the generation of insulin ions. It took at least 300 laser shots using relatively high power, 10 x P_t (sinapinic acid, insulin), to preheat the sample before the insulin ions could be observed. When a pure insulin sample without matrix was irradiated by the same laser using this high power, no insulin ions were observed. The "preheat process" may generate reaction products which are the functioning MALDI matrix instead of intact ninhydrin molecules. Figure 3.3 shows the MALDI spectrum resulting from a mixture of insulin and ninhydrin (molar ratio ~ 1: 1000) after 300 "preheat" laser shots were applied.

3. *9-Anthracenecarboxylic acid, N-phenylmaleimide, 2-naphthol, and 4-hydroxybenzaldehyde*

These compounds also were found to be successful MALDI matrices, yet the intensities of observed peaks for insulin ions were smaller. Figure 3.4 shows the MALDI spectrum resulting from a mixture of insulin and N-phenylmaleimide.

IV. Conclusions

Several chemical compounds were found to be successful MALDI matrices which have not been reported in the literature. However, the resulting quality of MALDI spectra was not good enough to give the newly discovered matrices advantages over commonly used matrices, such as sinapinic acid and α-cyano-4-hydroxycinnamic acid. These new matrices tend to generate analyte peaks of poor resolution or with strong matrix adducts, which may prevent practical use of the technique. However, the pursuit of understanding of how matrix molecules play a role in the MALDI process never ends. When more successful MALDI matrices are discovered, we will have a better chance to unveil the operating principles behind them.

The data presented here are preliminary. The experiments were done before we could appreciate what variables were most important and necessary to be documented. For example, the exact laser powers used to work on the successful MALDI matrices were not

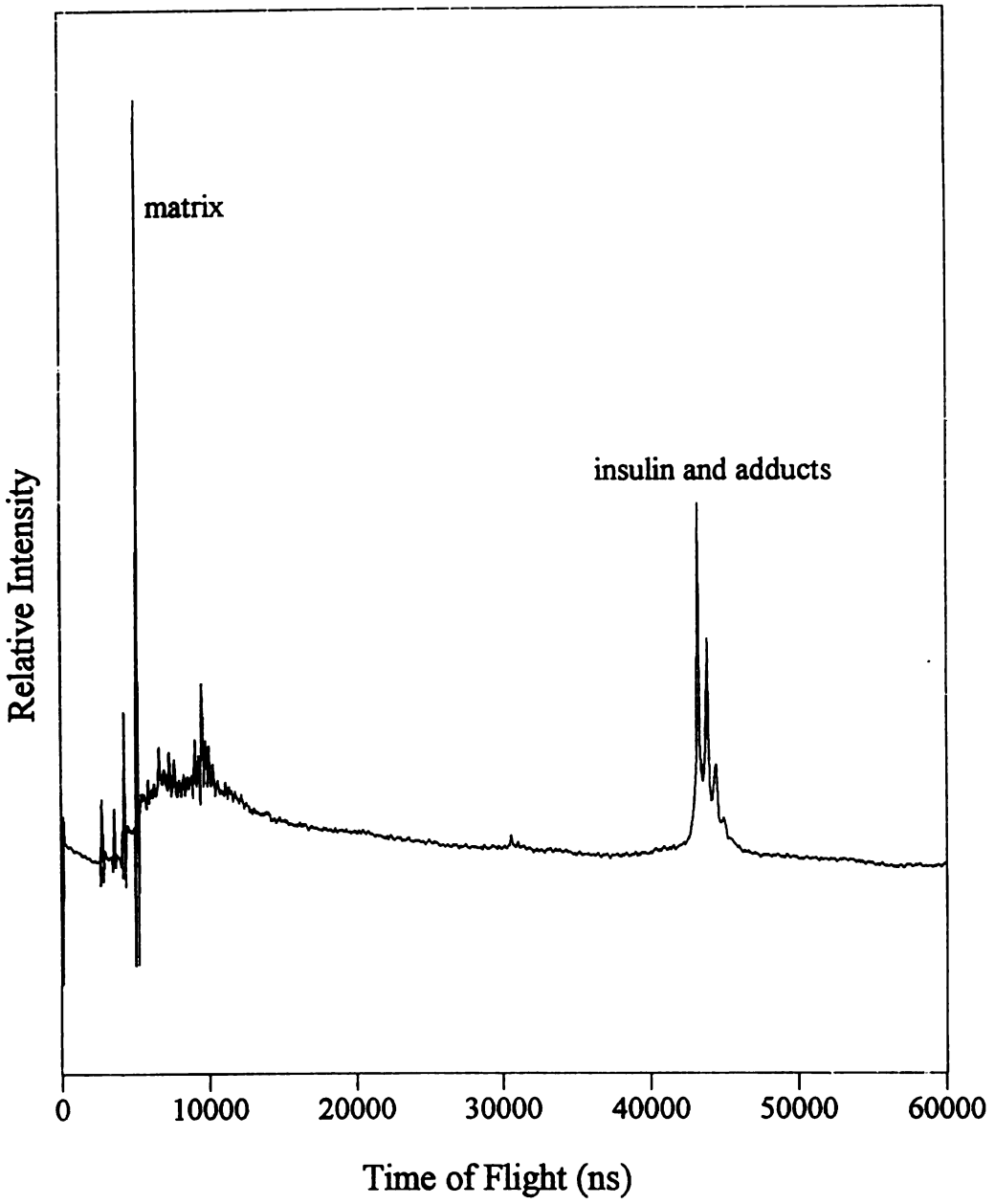


Figure 3.3 The MALDI spectrum resulting from the standard mixture of insulin and ninhydrin matrix (molar ratio ~ 1: 1000), after 300 "preheat" laser shots were applied.

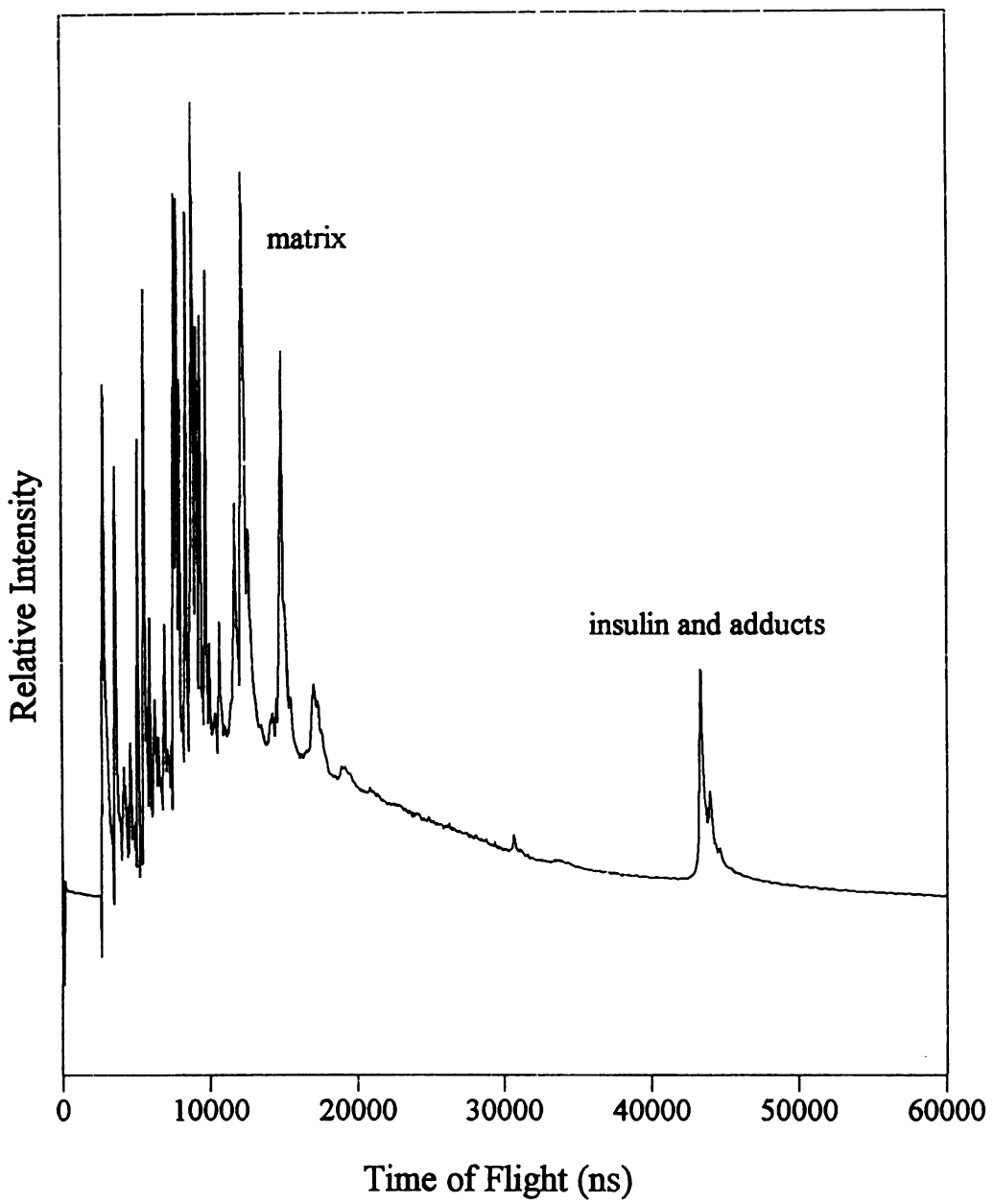


Figure 3.4 The MALDI spectrum resulting from the standard mixture of insulin and N-phenylmaleimide matrix (molar ratio ~ 1: 1000).

recorded. If this study were undertaken today, the "results" would be in quite a different form. Nevertheless, it is important to document these preliminary observations to assist those who are willing to pursue this topic in the future.

Chapter 4

Is There a Relationship Between Laser Power and Resolution?

I. Introduction

When MALDI-MS is applied to solve analytical problems, mass spectral resolution is a very important performance characteristic because it not only sets the limitation of accessible mass accuracy but also determines the ability to separate ions of similar mass in the spectrum. To date, time-of-flight (TOF)-MS is most commonly used for m/z analysis with MALDI. Several publications have discussed factors affecting signal resolution, and how to improve resolution in MALDI-TOF-MS [39-45]. A dominant factor that influences the mass spectral resolution in a MALDI-TOF experiment is the laser irradiance [46]. Laser irradiance must be above a threshold value for the MALDI process to occur. Laser irradiances much higher than the threshold value dramatically decrease the resolution in the spectra obtained. The peak widths of signals generated from a peptide, for example, are a function of the applied laser irradiances [44].

For simplicity, we will discuss "power" instead of "irradiance" in this report. Irradiance (power per unit area) is proportional to power if the beam profile remains the same. Most of the commercial MALDI instruments employ a nitrogen laser and a beam attenuator to vary laser power, however, the beam profile is not measured or controlled. A study using ultra-short laser pulses has shown that the amount of energy deposited into the sample, rather than laser power or irradiance, is the important parameter in the MALDI process (as long as the irradiation time is short) [28].

There is an unfortunate dilemma in MALDI-TOF-MS; both high intensity (high power) and high resolution (low power) cannot be achieved under the same conditions. There is also a power dependence to the MALDI process as a function of MW. The higher MW analytes require higher powers than do low MW analytes, for desorption/ionization (D/I) to occur (see Appendix I). When MALDI is used to investigate a mixture with a large MW range, it is necessary to use high laser power just to generate a signal at the high m/z values - at the expense of resolution throughout the spectrum.

Most MALDI spectra reported in the literature represent the summation or average of 20-200 mass spectral transients, which may be obtained from multiple locations on the MALDI target or obtained with a variety of power settings. Conclusions such as the high power-low resolution relationship seem to evolve from the analysis of such summed spectra, or from the spectrum resulting from a single laser shot. We recently showed that the mass spectral transients of MALDI-TOF-MS rapidly change with time, and an evaluation of the component spectra may yield different conclusions than evaluation of a single summed spectrum (The preprint of this publication is attached as Appendix III in this thesis). We use this tool of time-dependent MALDI to investigate details of the power-resolution relationship here. We will show here that, by changing how MALDI spectra are constructed, "low power resolution" can be realized in "high power" MALDI experiments.

II. Experimental

All MALDI results were obtained on a ResearchTec TOF mass spectrometer (Vestec Corp., Houston, TX), used in linear mode, equipped with a nitrogen laser (model VSL-337ND, Laser Science, Newton, MA, 337 nm, 3-ns pulse, 250 μJ/pulse). Laser light is attenuated using a variable attenuator (model 935-5-OPT, Newport, Fountain Valley, CA) with a computer-controlled stepper motor. That is, the laser power is software controlled

by setting the stepper motor position, corresponding to an on-screen number ranging from 0 to 2,400. A larger number results in a higher laser power. Laser powers are not expressed on an absolute scale, but relative to the value at the D/I threshold. The threshold laser power will be designated here as P_t (m, A), where P_t is the threshold power at which the analyte, A, generates observable signals (S/N > 5), using the matrix, m. To measure the laser power threshold, multiple samples were prepared using the procedure described below. Each fresh sample was irradiated consecutively, with laser power varied in increments of 100 units, until the analyte signal was observed. The threshold laser power P_t (αCN, I), where αCN represents α-cyano-4-hydroxycinnamic acid and I represents insulin, corresponds in our instrument to a stepper motor position of 1200. When the stepper motor position is set at 1800, the corresponding laser power is expressed in the text as 1.50 P_t. A laser spot size of ~ 0.8 mm^2 was used to irradiate sample on a 3.5 mm^2 target ("Big Spot MALDI") (see Appendix III). The nitrogen laser was prewarmed and then maintained firing for 15 minutes before any data were obtained; this has been found to decrease shot-to-shot variations. The accelerating voltage in the ion source was 26 kV. Data were acquired with a transient recorder with 2-ns resolution. All spectra shown represent unprocessed data. When resolution was measured, a Savitsky-Golay method was used to smooth raw data (second degree polynomial, 30 point smooth). The mass spectral resolution is defined by m/Δm (FWHM).

The matrix used was α-cyano-4-hydroxycinnamic acid. A water (with 0.1% trifluoroacetic acid)-acetonitrile (1:1, v/v) mixture was used as the solvent system to prepare a saturated matrix solution (~ 70 mM) and a stock solution containing both bradykinin and bovine insulin (10 pmol/μL for each analyte). A sample solution was prepared by diluting the peptide stock solution with the matrix solution to give a final concentration of 2 pmol/μL for each analyte. To prepare samples for MALDI analysis, 2 μL of the sample solution were applied to a flat stainless steel probe tip. The sample mixture was then allowed to air dry, leaving 110 nmol of matrix, 4 pmol bradykinin, and

4 pmol insulin on the target, before introduction into the ion source. Time-to-mass conversion was achieved by either external or internal calibration using peaks for Na^+ and K^+, most abundant matrix peaks, and peaks from bradykinin and insulin.

Bradykinin and bovine insulin were purchased from Sigma Chemical Co. (St. Louis, MO). α-Cyano-4-hydroxycinnamic acid was purchased from Aldrich Chemical Co. (Milwaukee, WI). Acetonitrile and trifluoroacetic acid were purchased from EM Science (Gibbstown, NJ). The compounds were used as purchased without further purification.

III. Results and Discussion

If we use the typical approach of evaluating summed data, results shown in Figure 4.1 are obtained. Figure 4.1a and 4.1b show the mass spectral peaks representing insulin ions (presumably, protonated insulin molecules) obtained at two different laser powers; both represent the sum of 300 transients. The peak obtained using higher laser power exhibits lower resolution. From such spectra obtained at a variety of laser powers, resolution can be calculated. The results are summarized in Figure 4.1c. The resolution decreases considerably with increasing laser power for both bradykinin and insulin signals. This observation parallels that reported by Ingendoh *et al.* [44].

Figure 4.2 shows the MALDI-TOF mass spectrum obtained at a laser power of 1.33 P_t (αCN, I) by averaging 500 transients. Figure 4.3 shows a series of component spectra obtained by dissecting the spectrum shown in Figure 4.2. To generate these spectra, the experiment was configured so the laser fires 10 times in 2.5 seconds. The laser then stops while the transient recorder downloads the file representing the average of these ten (a "component spectrum"), and a new file representing the next 10 mass spectral transients is generated (see Appendix III). Each single component spectrum in Figure 4.3 is the sum of 10 sequential mass spectral transients. Figure 4.3 contains 50 component

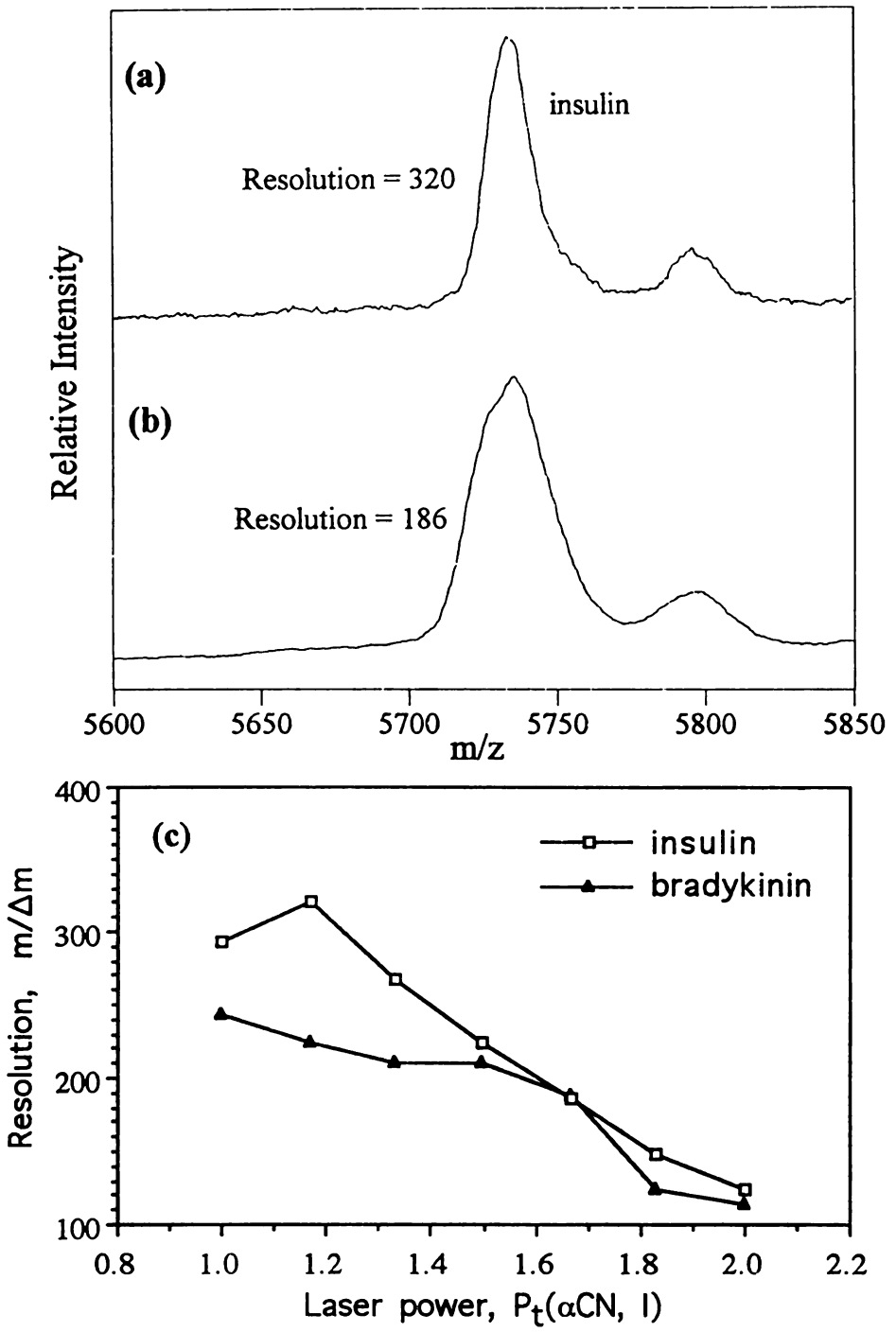


Figure 4.1 The mass spectral peaks representing insulin ions (presumably, protonated insulin molecules) at two different laser powers (a) 1.33 Pt (αCN, I) and (b) 1.67 Pt (αCN, I); (c) resolution vs. laser power is plotted; all data represent the sum of 300 transients. 4 pmol of bradykinin, 4 pmol of insulin, and 110 nmol of matrix were used to prepare each MALDI sample.

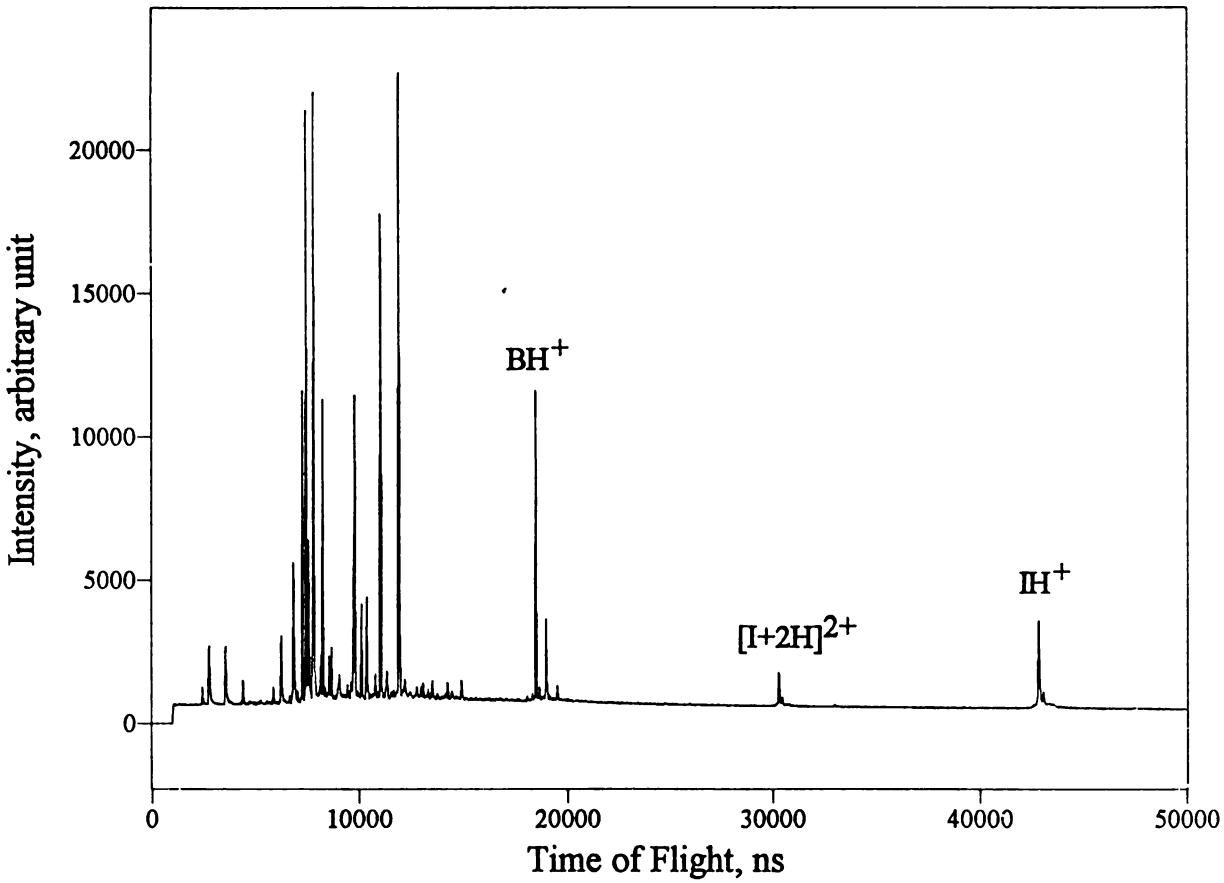


Figure 4.2 The MALDI-TOF mass spectrum obtained at a laser power of 1.33 Pt (αCN, I) by averaging 500 transients. 4 pmol of bradykinin, 4 pmol of insulin, and 110 nmol of matrix were used to prepare the MALDI sample. Abbreviations used: B, bradykinin; I, insulin.

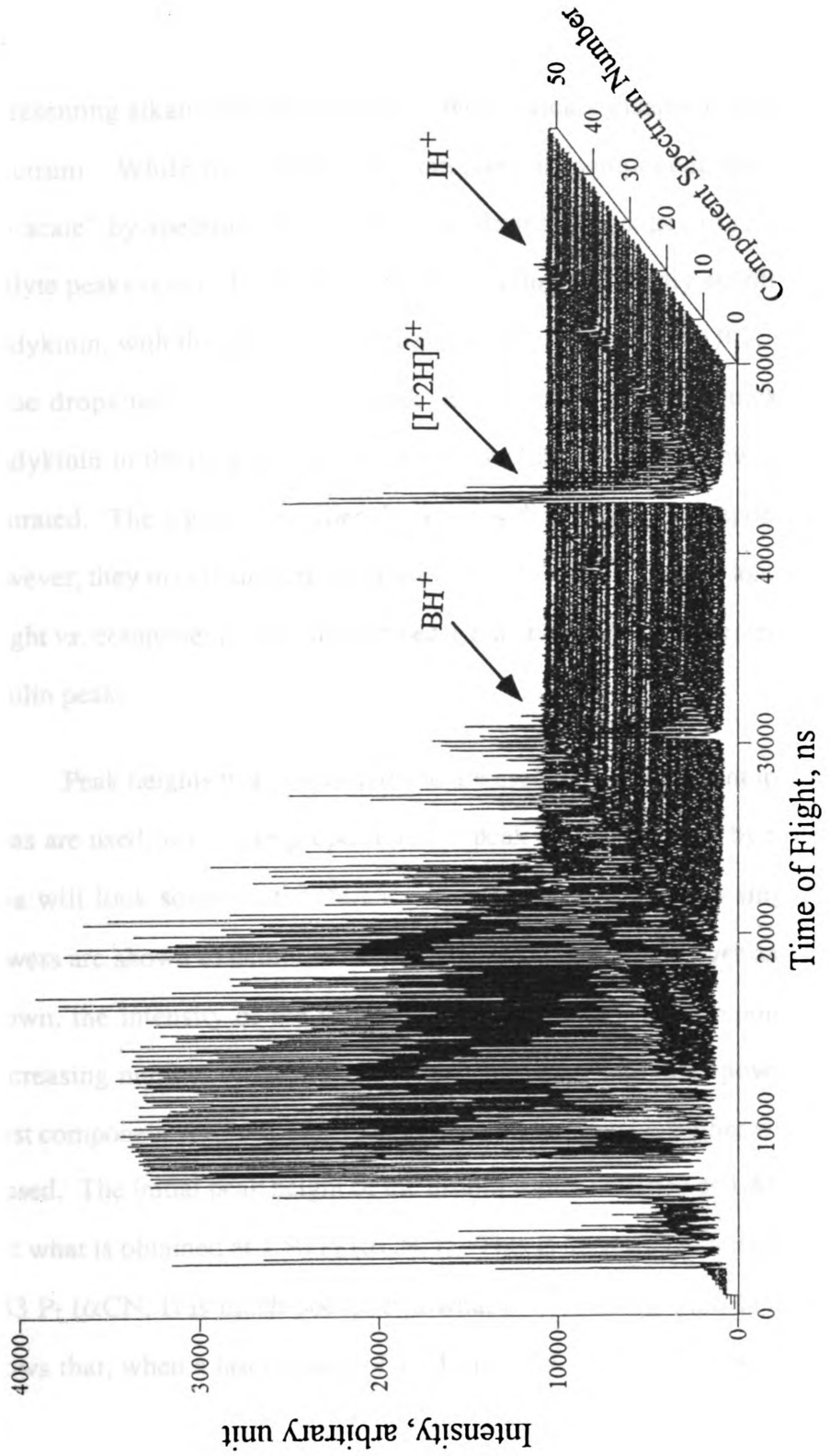


Figure 4.3 A series of component spectra obtained by dissecting the spectrum shown in Figure 4.2. Each single component spectrum is the sum of 10 sequential mass spectral transients. Abbreviations used: B, bradykinin; I, insulin.

spectra. A few component spectra are shown in Figure 4.4. Consider Figure 4.4a and 4.4b, which show component spectra #2 and #12. In the earliest spectrum, #2, peaks representing alkali ions are prominent; these quickly disappear by the twelfth component spectrum. While many peaks are saturated in component spectrum #2, they are all "on scale" by spectrum #12. Also note, dramatic changes in relative intensities of the analyte peaks occur. In spectrum #2, the insulin peak is only slightly smaller than that for bradykinin, with the ratio of intensities for IH^+: BH^+ equal to 0.7. By spectrum #12, this value drops to 0.2; the insulin peak decreases much more quickly than does that for bradykinin in the data set. As the data show, many peaks at low m/z values are initially saturated. The signal intensities decrease with increasing component spectrum number, however, they may then increase at some later time and decrease again. The changing peak height *vs.* component spectrum number is quantitatively documented in Figure 4.5a for the insulin peak.

Peak heights were measured and are used here to represent ion abundances. If peak areas are used, which are proportional to peak heights divided by resolution, then Figure 4.5a will look somewhat different. Data from experiments using three different laser powers are shown to demonstrate the relationship of laser power and signal intensity. As shown, the intensity of the insulin signal decreases with component spectrum number (increasing number of laser shots), regardless of which laser power is used. The initial (first component spectrum) intensity of the insulin signal is higher when higher laser power is used. The initial peak height of the insulin signal obtained at 1.83 P_t (αCN, I) is smaller that what is obtained at 1.50 P_t (αCN, I). This is because the initial resolution obtained at 1.83 P_t (αCN, I) is much poorer that what is obtained at 1.50 P_t (αCN, I). Figure 4.5a shows that, when a laser power of 1.33 P_t (αCN, I) is used, about 20 useful spectra with reasonable intensities (for the insulin peak) are generated. When a laser power of 1.80 P_t (αCN, I) is used, more than 60 useful spectra with reasonable analyte peak intensities are generated. So the signal lasts longer when higher laser powers are used.

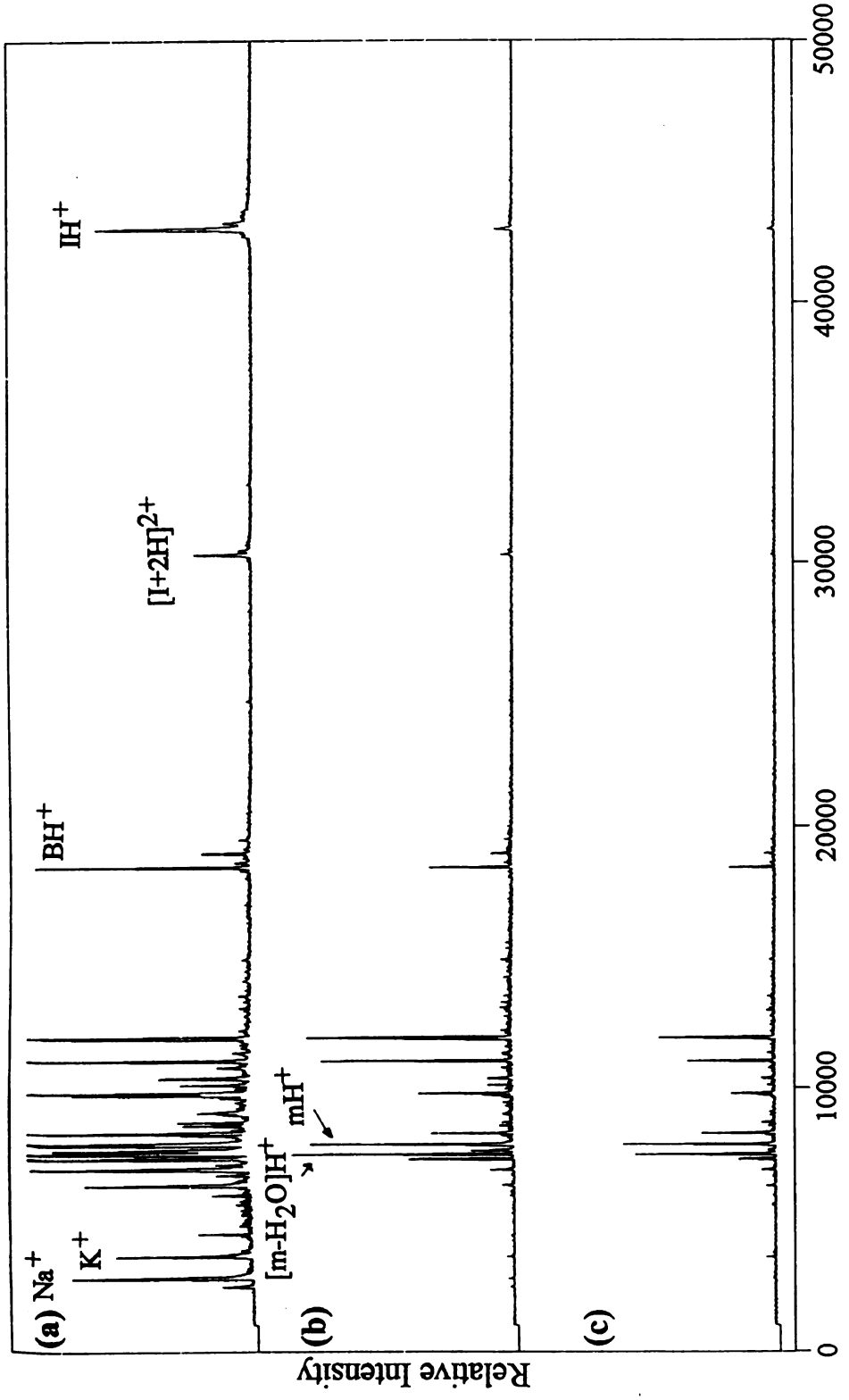


Figure 4.4 Three component spectra from Figure 4.3; component spectra (a) #2, (b) #12, and (c) #45. Abbreviations used: B, bradykinin; I, insulin; m, α-cyano-4-hydroxycinnamic acid.

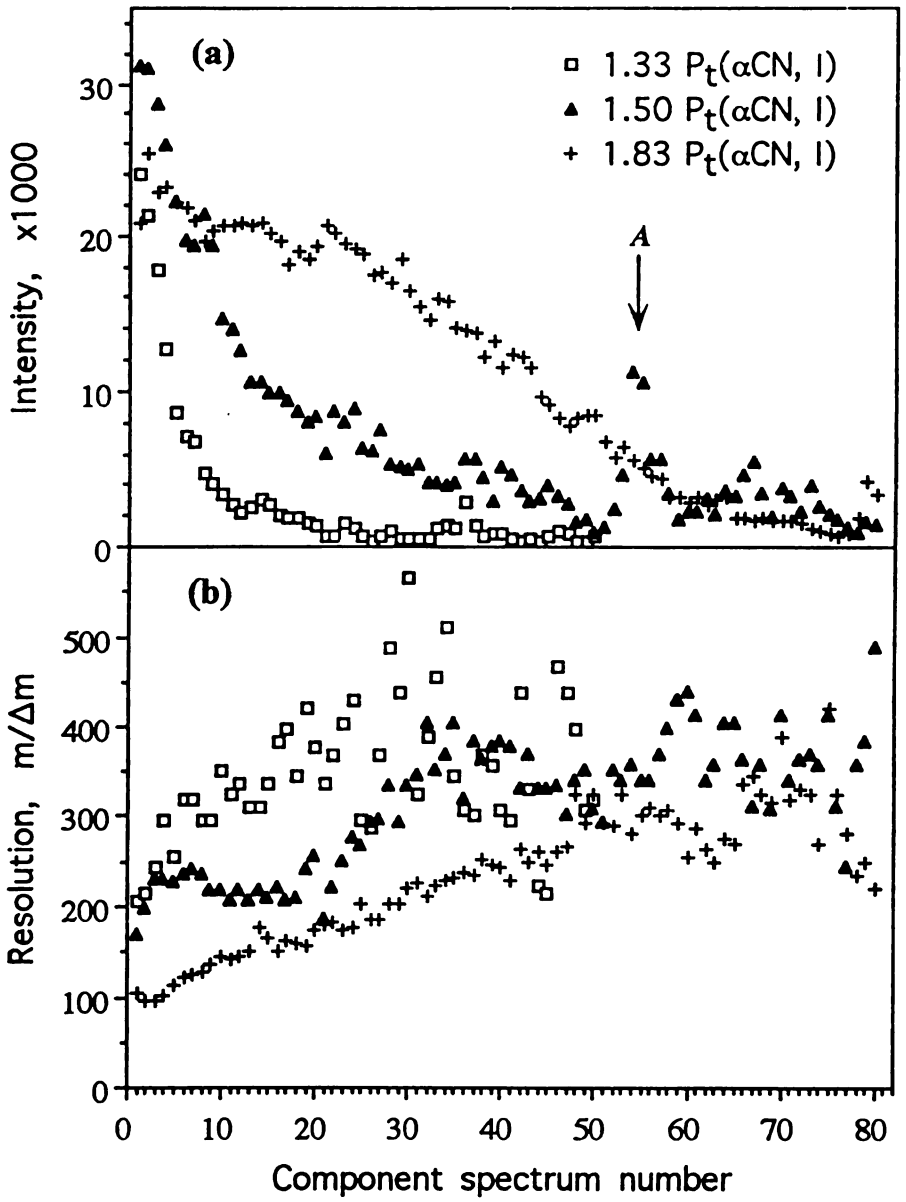


Figure 4.5 The (a) peak height and (b) resolution of the insulin signal obtained at different laser powers are plotted against component spectrum number.

The total ion current (TIC), which is proportional to the summation of insulin peak intensities over all the component spectra (area under the curves in Figure 4.5a), is also larger when higher laser power is used. The data in Figure 4.5a correlate with experience - signals "last longer" and are more intense at higher powers.

The resolution in each component spectrum can also be measured. It is found that the resolution in all of the component spectra is not constant. In Figure 4.5b, insulin signal resolution is plotted against component spectrum number. When the laser power is 1.33 P_t (αCN, I), the data show the highest resolution. The resolution begins at approximately 200 and increases with component spectrum number to 300 ~ 400, but considerable scatter of measured resolution appears when the component spectrum number is larger than 20. This kind of scatter of measured resolution is often observed when the peak intensities are small. The initial resolution of the insulin signal is lower when higher laser power is used. When the laser power is 1.50 or 1.83 P_t (αCN, I), the data show that the resolution increases with component spectrum number. When the laser power is 1.80 P_t (αCN, I), the measured resolution starts at ~ 100 and increases to ~ 300, while the intensity falls with increasing component spectrum number. That is, the "low power resolution" (300 ~ 400) can be realized when "high power" is used.

Experimental Implications

When the data of MALDI-TOF-MS are evaluated in the form presented here, one may envisage some variations in how the experiment is done, and possibly different approaches for obtaining the most useful data from an experiment. For example, when we used a smaller, focused laser spot (as our commercial instrument was supplied), we would frequently move the spot around so that most of the transients would contain strong signals. We now see that we are collecting transients with the lowest resolution in this way, and there are advantages to using a larger fixed spot, and accumulating more

transients. However, if it is not resolution but signal intensity that is most important, then moving the laser while summing is a good approach.

With a set of spectra representing the experiment, one can evaluate spectra and decide which may be summed to yield a single spectrum, of higher quality, representing the MALDI analysis of a compound or mixture. For example, Figure 4.6 shows the peak shapes and corresponding resolution of insulin signals, from summing different ranges of component spectra which are obtained at 1.50 P_t (αCN, I). Since the early laser shots yield poor resolution, one can take advantage of this information to improve the resolution in the final spectrum. By rejecting the spectral transients from early laser shots, one can easily improve resolution when only the best transients are summed. Using Figure 4.5b, one can select an appropriate range of component spectra for summation to yield resolution improvement. In Figure 4.5, the data obtained at 1.50 P_t (αCN, I) suggest that summation of component spectra 31-40 may produce a quality spectrum of good resolution and reasonable signal intensity. As shown in Figure 4.6, the resolution of the insulin signal is improved from (b) 220 to (c) 370.

A third implication of the data may be related to mixture analysis, and this is an approach that we will be evaluating in subsequent publications. Suppose a mixture such as an enzymatic digest was being analyzed by MALDI. We need to know how many peptide fragments were formed, and the best possible mass assignment for each. One approach is to obtain a series of spectra at very high laser powers. High power would optimize our chances to detect those components with the smallest desorption/ionization efficiencies, and could be least hindered by the power/molecular weight dependence. In subsequent spectra, as the intensity of each peak decreases, the user can decide when (i.e., in which component spectrum) the mass of the analyte will be determined. All peaks need not be analyzed in a single spectrum - we can select which spectrum is appropriate for each, when the resolution is sufficiently high that the best result for that instrument can be realized. This would be

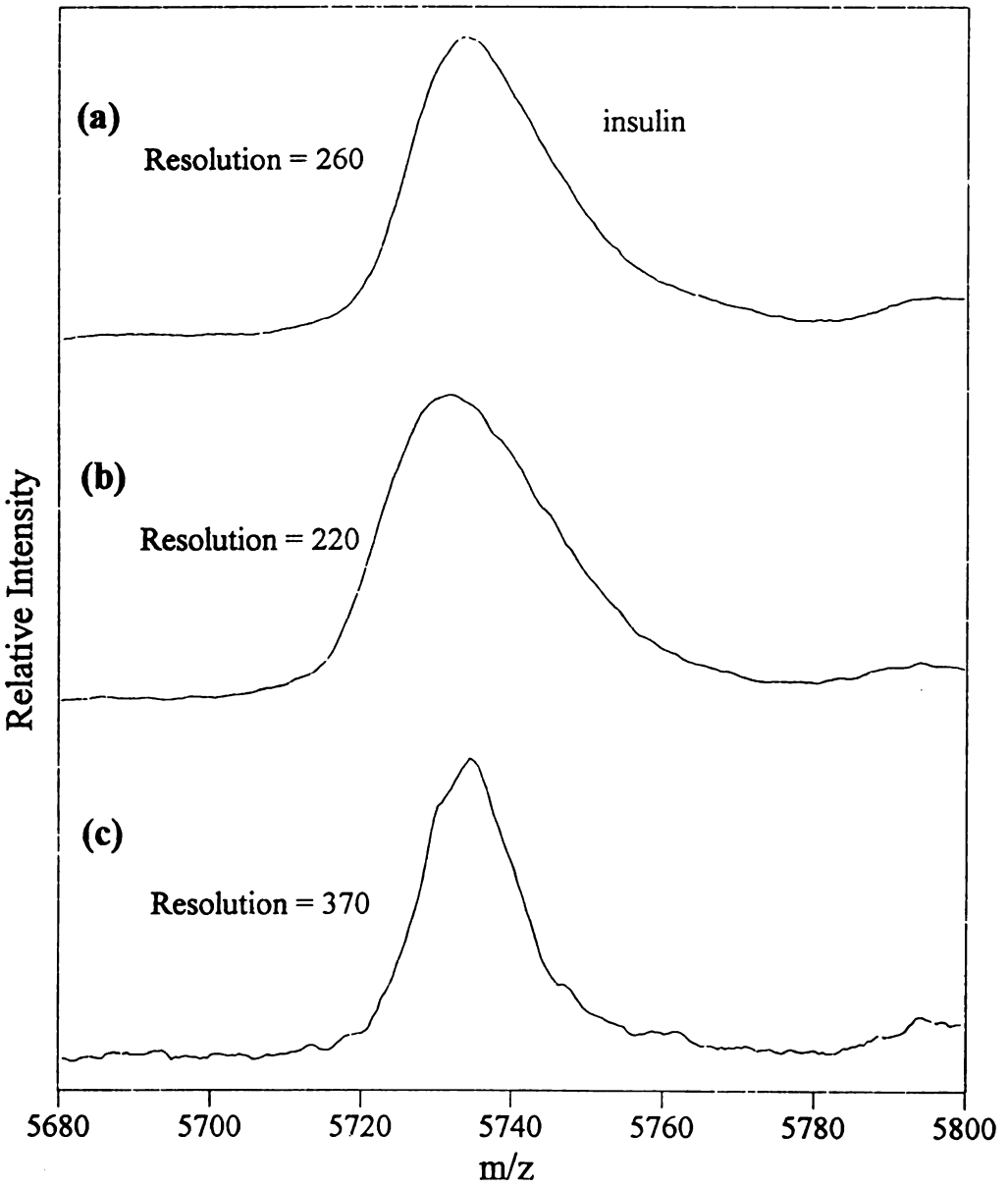


Figure 4.6 The peak shapes and corresponding resolution of insulin signals, from summing different ranges of component spectra obtained at 1.50 Pt (αCN, I), (a) #1-#80, (b) #1-#10, and (c) #31-#40.

very different from the typical approach that we believe most labs use, which involves analysis of data contained in a single mass spectrum from the MALDI-MS analysis.

Changing Resolution at a Single Power

While we feel that it is important to document the fact that a variety of resolutions can be realized, even at high power, it is not the purpose of this work to provide an explanation. Definitive experiments remain. A key problem at this point is in the description of the observation. We can state that, because of the way the data were collected/presented, Figure 4.5b shows that resolution increases as the component spectrum number, i.e., number of laser shots, increases. This is not to suggest that there is a cause-and-effect relationship between these two quantities. By considering the data presented in Figures 4.4a and 4.4b, another possible description of the observation is that resolution increases as peak intensity decreases. Either could be the actual situation. If the total amount of material ablated per laser shot is largest in the first shot of the experiment, and resolution depends on various thermal and collisional aspects of the desorption plume, then one might well expect resolution to depend on the how many laser shots had been fired at a given "spot". Alternatively, we know that we can be generating a high ion density in this experiment, and ion-ion repulsions because of ion densities exceeding the space charge limit would certainly affect the spatial aspects of isomass ion packets as they traverse the instrument, resulting in broader peaks for high ion abundances (intense peaks). As Figure 4.5a suggests, there is a definite relationship between intensity and accumulated irradiation time, which makes it difficult to make a statement precisely describing the variable that best correlates with resolution. However, it is useful to make a single plot collecting the results of many experiments involving different powers, into a single intensity *vs.* resolution plot, which is shown in Figure 4.7 for insulin ions. This reinforces the observation that the complete range of resolution can be realized by working at high power - the resolution in early spectra is low, but constantly improves as irradiation

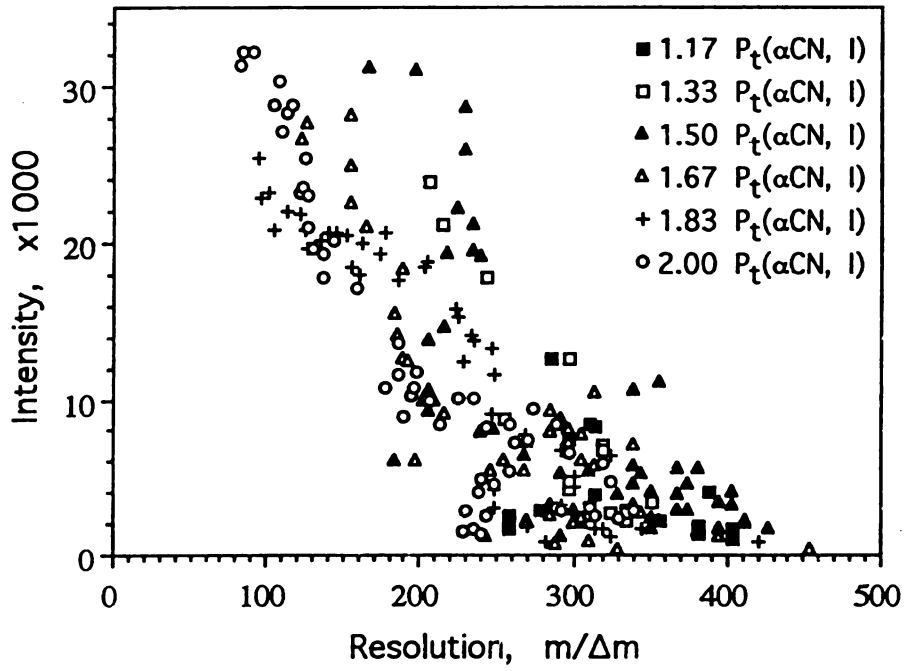


Figure 4.7 Intensity vs. resolution plot, for peaks representing insulin ions, obtained at different laser powers.

continues. While there is obviously scatter in the data, there is a general trend in the data as plotted in Figure 4.7, where larger peaks yield the lowest resolution.

We cannot, here, reach a conclusion on which variable of the experiment is determining the resolution. While a resolution/intensity correlation is attractive for the reasons mentioned above, details of the data presented may not support this selection. The data in Figure 4.3 show that the insulin peak is initially large, drops off, but grows again in component spectrum # 36. This feature is indicated as point A in Figure 4.5a. If there is a resolution/intensity relationship that is operational, this increase in insulin ion intensity in 4a should be accompanied by a dip in the resolution data at that point in 4b. This is not observed. Thus, a precise description of the observation, correctly indicating that resolution is a function of some experimental variable, remains in this preliminary report, elusive.

Can the sample be characterized as having a distribution of thresholds?

While the data in Figure 4.5 is a quantitative representation of an aspect of MALDI for which practitioners are aware, is it expected? Consider the data for the lowest power, 1.33 Pt, in Figure 4.5a. The intensity starts at 24,000 and approaches zero by the 20th component spectrum. If this were GC/MS , for example, the peak would be gone when the sample concentration is zero. That is, after 200 laser pulses the sample (insulin) seems to be depleted. This is obviously not correct. As data for other power settings in Figure 4.5a show, one can get many more insulin ions from the same location at higher power. The initial intensities are higher at higher power, and the signal lasts longer. We also know that, if we let the signal go to zero and then raise the power, additional analyte ions can be formed. We understand that, when power is increased, the effective size of the laser spot can increase - if the power is not constant across the laser spot. However, we do not believe that the observations can be explained based on such changes alone. We

believe that the data suggest that some threshold distribution exist, or is formed, within the sample target.

Again, the challenge here is how to describe this observation. We will propose that, in addition to the specific selections of the matrix and analyte, the MALDI threshold also depends on a variable, x, which describes the threshold distribution. That is, $P_t = f(m, A, x)$. This has been described in other ways in the many facets of the technique that have been discussed and considered to date.

One possible physical aspect of the MALDI matrix that could be responsible for the threshold distribution is the range of sizes of the crystals formed on the surface - that is, $x = s$, as shown in Figure 4.8a. The size of the crystals prepared in our experiment is certainly not homogeneous. If one considers models for the desorption process in MALDI, the first step is deposition of energy due to matrix absorption. The temperature in some portion of the matrix crystal rapidly increases, although energy rapidly flows away from that portion of the crystal as well. If the crystal is sufficiently small, and heat dissipation mechanisms are limited, high temperatures and prompt ablation is realized. For large crystals, only a portion of the crystal may be ablated per shot, or higher powers may be required, since heat can be conducted throughout the crystal, away from the point of excitation. Thus, smaller crystals may have lower MALDI thresholds than larger crystals. As the power is increased, the range of crystal sizes over which irradiation leads to signals increases.

The successive ablation of material with every laser shot in MALDI has also been discussed in terms of depth profiling, leading to a second possibility where $x = d$, the distance below the surface, into the crystalline target, as suggested in Figure 4.8b. Since many thousands of crystals can be formed on the small MALDI target, a variety of mechanisms may exist which would lead to a depth-dependent threshold. This may be related to "crystal packing" aspects of the target, or due to target heterogeneity, since one

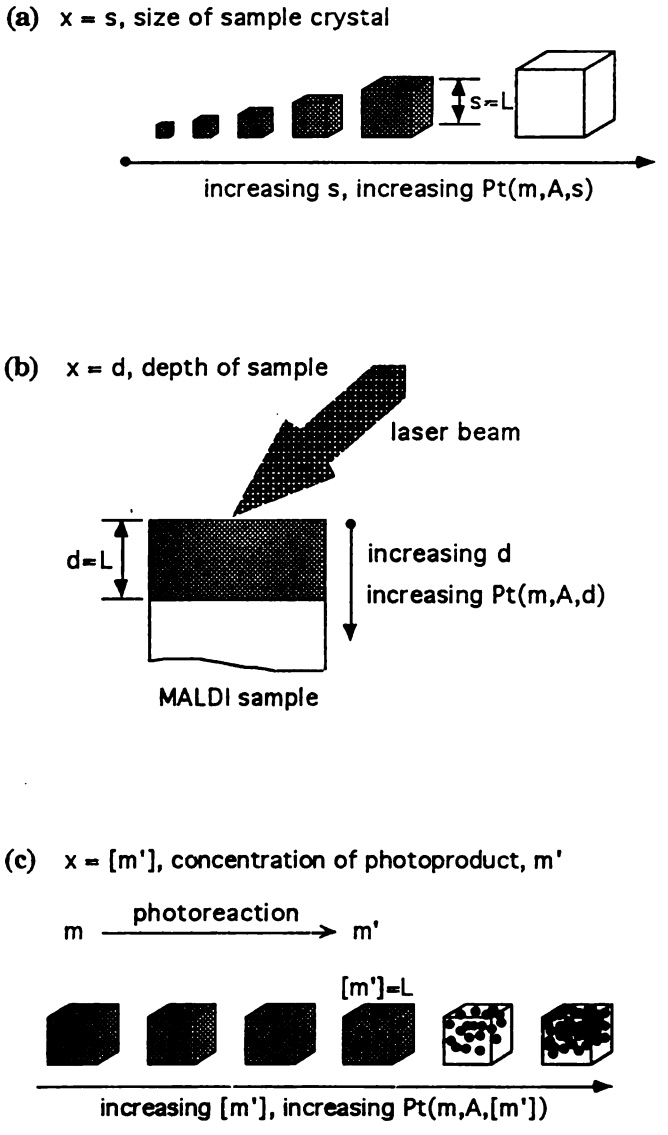


Figure 4.8 Physical aspects of the MALDI matrix that could be responsible for the threshold distribution. In addition to the specific selections of the matrix and analyte, the MALDI threshold also depends on a variable, x, which describes the threshold distribution. That is, $P_t = f(m, A, x)$. (a) x = s, the size of the crystals, (b) x = d, the distance below the surface, into the crystalline target, and (c) x = [m'], concentration of a photoreaction product of the matrix, m, where m' is not a "good" matrix. In (a), (b), and (c), when a fixed laser power is used, only the samples which fulfill the condition (a) $s < L$ (limit) (b) $d < L$ (c) $[m'] < L$ can be desorbed.

may expect some sort of "layering" within the crystals. The compounds present do have very different solubilities. It is not hard to see that, as the laser continues to fire, material is removed in each laser shot. When some initial power is chosen, we may eventually achieve a depth, d = L (limit), at which the threshold becomes higher than the power selected, and no additional signal is generated unless the power is increased.

A third possibility may be that x is not a dimension but a time-related variable, equal to the total irradiation time ($t = n_{pulses} \times 3$ ns/pulse). It has been suggested that a variety of photochemical processes can occur in the MALDI target. Suppose that some fraction of the matrix is converted into a new compound with each laser pulse, and the compound has a lower molar absorptivity and higher heat of sublimation. It does not vaporize as readily as does the original matrix molecules, and tends to accumulate as t increases. This photoproduct accumulation causes the threshold to be detectably higher, after n shots (Figure 4.8c).

A final possibility could, of course, be a combination of mechanisms such as those described here. The point to realize here is that the experiment behaves as though some threshold distribution exists, as suggested by Figure 4.9 and one can always get more signal by going to some higher power, at least throughout the range of powers accessible in this experiment. It should be made clear that, in this discussion, we are evaluating $P_t(m, \underline{A}, x)$ - the threshold distribution for generation of ions from a specific analyte. We are not suggesting that, when the insulin peak approaches zero, all of the other peaks disappear as well. For example, at a power of 1.33 P_t, the insulin peak has effectively gone to zero by component spectrum # 45. This spectrum is shown in Figure 4.5c. As the figure shows, while the insulin peak is very small, the bradykinin and many matrix peaks are still present. They have different P_t values and P_t distributions; this aspect must be considered as well when attempting to describe features of the experiment that lead to the threshold distribution.

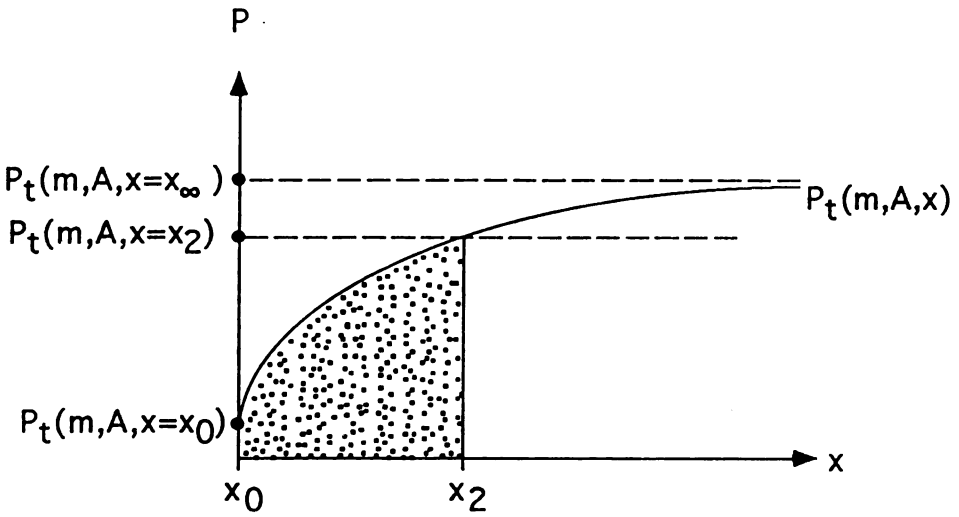


Figure 4.9 The conceptual threshold distribution function, Pt (m, A, x), on a P-x plane, where P is applied laser power and x is a variable which describes the threshold distribution. When a laser power equivalent to Pt (m, A, x = x1) is used, only a portion of the sample which meet the requirement, Pt (m, A, x) < P = Pt (m, A, x = x1), the dotted area in Figure 8a, can be desorbed/ionized.

IV. Conclus

We

interesting

capabilities

large numb

single spec

magnet and

source.

systems th

found rea

through t

We began

MALDI s

realized.

spectra c

We show

complex

powers

analysis

from the

from a

We con

compon

continu

IV. Conclusions

We find the parallels between the development of MALDI and that of GC/MS interesting to compare. In many laboratories, gas chromatographic sample introduction capabilities were available on mass spectrometers before data systems, that could collect large numbers of spectra, were available. One would frequently have to be content with a single spectrum per mixture component, obtained by manually triggering the scan of the magnet and, simultaneously, an oscillographic recorder, as a component eluted into the ion source. Conclusions were drawn from this single spectrum. Eventually, with data systems that could record spectra as fast as the mass spectrometer could generate them, we found real advantages in characterizing the time-dependence of mass spectral peaks (through the use of "mass chromatograms"). A similar evolution can be seen in MALDI. We began by reporting single spectra, frequently the sum of many transients, as THE MALDI spectrum - on which analytical results, and characteristics of the technique, were realized. We now see that MALDI spectra change with time, and the time-dependent spectra can provide more information than a single spectrum representing the experiment. We show here that the relationship between MALDI resolution and laser power is more complex than previously realized, due to the time dependence of the resolution. High powers can be used and, while the resolution may be low in spectra obtained early in an analysis, later spectra exhibit higher resolution. To fully realize this, and be able to benefit from the effect, larger laser spots should be used. This leads to a larger number of spectra from a "single spot", without having to move the laser to different locations on the target. We continue to find that the ability to "dissect MALDI spectra", to collect and evaluate component spectra as well as construct their summed result, is a useful tool in the continuing effort to define and develop MALDI-TOF-MS.

Chapter 5 - An Approach to Locate Phosphorylation Sites in a Phosphoprotein: Mass-Mapping by Combining Specific Enzymatic Degradation with Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry

This work has been published and the reprint is attached as Appendix IV. The phosphorylation sites of the protein Op18, in transformed lymphocytes, have been investigated by this method and the reprint of the resulting publication is included as Appendix V. The details of the computer program written to perform mass mapping were not covered by these publications and are discussed in this chapter. A further refinement of the experimental procedures for the enzymatic dephosphorylation reaction, a critical step in the location of phosphorylation sites using MALDI, is also discussed.

I. Mass mapping program

A computer program, named as MSU MassMap, was written using Microsoft® Visual Basic™ version 3.0 for Windows™ (Microsoft Corporation, Redmond, WA) to calculate the masses of possible peptide fragments and phosphopeptide fragments from the specific enzymatic or chemical degradation of a protein, and the m/z value of the mass spectral peak for the corresponding $[M+H]^+$ ion. Because of its limited resolution, MALDI-TOF mass spectrometry provides only the average masses of ions rather than their monoisotopic masses when the isotopic peaks are not resolvable in the high mass range. In the program, the average masses were used instead of monoisotopic masses which are typically used by programs designed for MS analysis (e.g., MacProMass software, by Terry Lee and Sunil Vemuri, Beckman Research Institute of the City of Hope, Duarte,

CA). Because partial digestion is not an uncommon situation even when long incubation times are used, the program computes masses for peptides from not only complete, but also partial digestion.

The source code of this program is appended to this thesis as Appendix VI. An instruction protocol for the use of MSU MassMap is included in Appendix VII. The inputs, operations, and outputs of this program are summarized below. Figure 5.1 shows a diagram of the algorithm using peptide, KRPSQRHGSKY, as an example. All the descriptions below use this diagram. The inputs require the following:

(1) Protein sequence: A text file is built to contain the name of the protein and its amino acid sequence. The file format is defined in Appendix VII. In the example of Figure 5.1, the peptide has the sequence, KRPSQRHGSKY.

(2) N-terminal or C-terminal modification of this protein: The mass of the N-terminal or C-terminal modification (or both) is defined by user interfaces. An user interface is an interactive device which allows users to exchange information with the program during run time. In the example of Figure 5.1, there is no terminal modification.

(3) The rules of specific degradation: These are defined by user interfaces. In the example of Figure 5.1, trypsin cleaves the amide bonds after (at C-terminal of) arginine and lysine residues.

(4) The residues which possibly undergo phosphorylation: These can be none or any combination from serine, threonine, and tyrosine residues, which are defined by user interfaces, too. In the example of Figure 5.1, they are serine, threonine, and tyrosine.

(5) Mass range: Only the peptide fragments with masses within this range will be considered for calculations. This is defined by user interfaces.

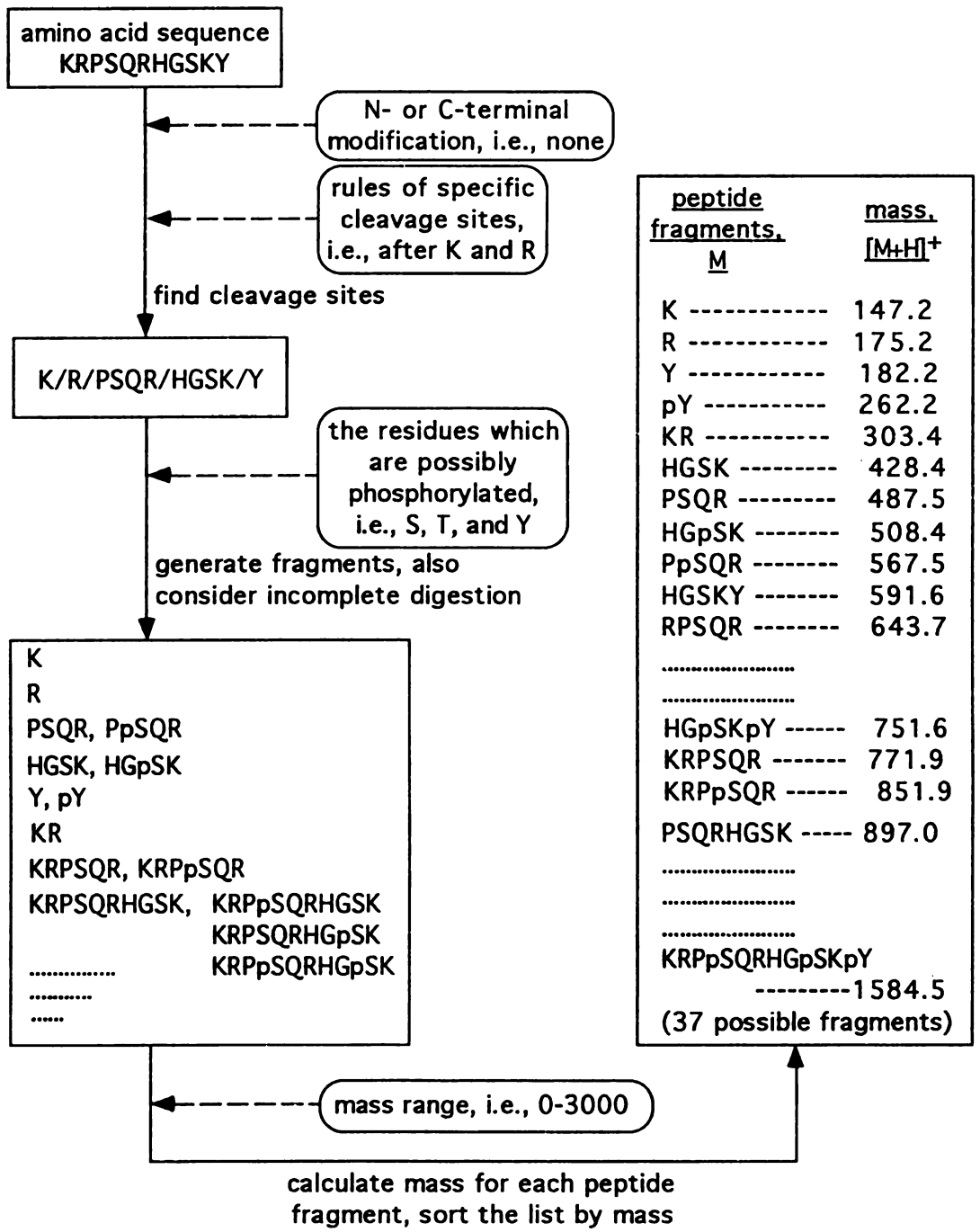


Figure 5.1 The diagram shows the algorithm of the mass mapping program using trypsin on the peptide, KRPSQRHGSKY, as an example. A lower case p preceding the phosphorylated residue indicates that the peptide is phosphorylated at that residue.

The program then performs the following operations:

(1) Generate a list which contains all possible peptide fragments, by following the rules of specific degradation from complete and partial digestion.

(2) Attach the possible phosphorylation sites to the peptide fragments generated in the above list. In the example of Figure 5.1, a peptide fragment, KRPSQRHGSK, which contains two serine residues would propose two more possibilities: mono- and di-phosphorylated peptides.

(3) Calculate the mass for each peptide fragment (includes phosphopeptide) in the list. The mass of the $[M+H]^+$ ion is calculated, as shown in Figure 5.1.

(4) Sort the peptide fragments in the list by their masses.

The output is a printout or a file which contains:

(1) All information from inputs.

(2) A list contains all possible peptide fragments, sorted by their mass, considering complete/partial digestion and possible phosphorylations. In the example of Figure 5.1, there are 37 possible fragments. Each peptide fragment in the list is defined by its assigned start/end amino acid residue numbers and number of phosphorylations on it.

An example of program output is included in Appendix VIII.

II. Enzymatic dephosphorylation reaction of immobilized phosphopeptides for direct analysis by MALDI

Phosphopeptides are identified by -80 (or multiple of -80)-Daltons mass shifts in the mass spectra after dephosphorylation with alkaline phosphatase. To minimize sample loss and facilitate easy, fast analysis of this method, especially when only picomolar amount of samples are available, it is advantageous to reduce sample manipulations. A

protocol has been designed to immobilize phosphopeptides on a supporting material, wash sample when necessary, conduct enzymatic dephosphorylation, remove reaction buffers, add matrix compounds, then analyze the final reaction products by MALDI *in situ*. The material must meet both requirements for efficient separations and an effective interface to a MALDI mass spectrometer. Nylon-based membranes have been reported to exhibit these properties [47]. The use of zetabind (0.45-μm pore size, 50-μm thickness, Cuno Laboratory Products, purchased from Life Science Products Inc., Denver, CO), a positive charge-modified nylon membrane, is described here for a demonstration.

A diagram in Figure 5.2 shows an outline of the protocol. 10 pmol of phosphopeptide, KRPpSQRHGSKY-amide, was immobilized on a piece of membrane fixed on a stainless steel probe by air-drying the peptide solution. The sample probe was washed with 0.5 mL 0.1% trifluoroacetic acid solution twice to remove any water-soluble contamination. 2 μl of 50 mM NH_4HCO_3 buffer solution (pH 8.0) containing about 3 units of calf intestine alkaline phosphatase (Boehringer Mannheim Biochemicals, Indianapolis, IN) was applied to the sample probe. An aluminum block was made to hold sample probes, and a glass chamber with air-tight cap was made to hold the aluminum block, so the whole assembly can be incubated in a water bath at 37°C. Extra 50 mM NH_4HCO_3 buffer solution was added to the chamber to establish the vapor pressure, so the reaction buffer solution on the probe will not dry out. After 4 hours of reaction time, the sample probe was washed with 0.5 mL 0.1% trifluoroacetic acid solution twice to remove reaction buffers. Then, 1 μl of 20 mM α-cyano-4-hydroxycinnamic acid matrix solution (in 0.1% trifluoroacetic acid solution: acetonitrile = 1:1) was added to the sample probe. The sample probe was covered to allow slow drying to help the elution of peptides from membrane surface to matrix solution. When the peptide/matrix mixture was dried and cocrystallized on the membrane, the sample was subjected to analysis by MALDI. Figure 5.3 shows the expected -80 Daltons mass shifts in the MALDI spectra due to the loss of phosphate moieties after using this protocol.

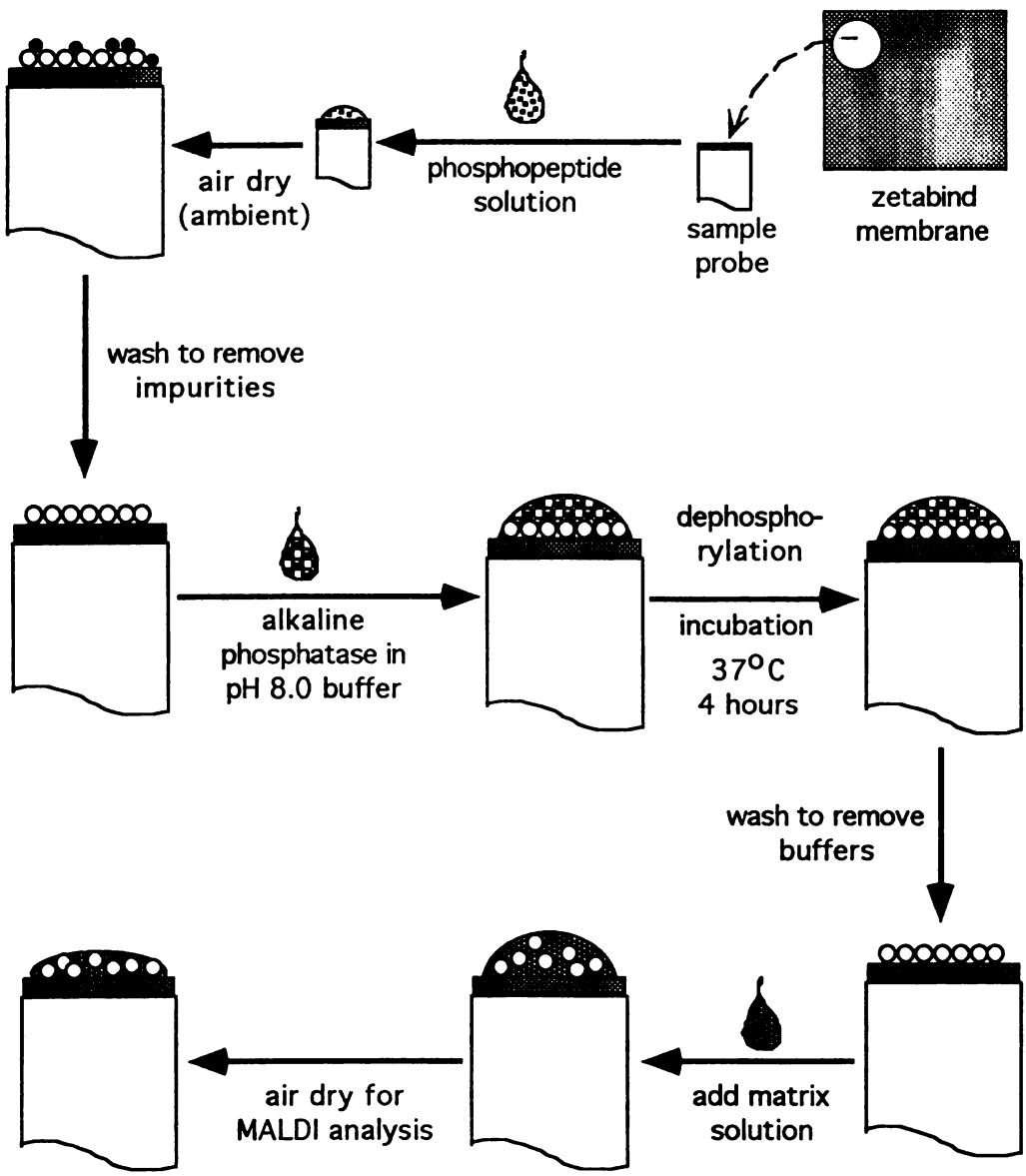


Figure 5.2 Sample immobilization and washing protocol for enzymatic dephosphorylation reaction of phosphopeptides for direct analysis by MALDI.

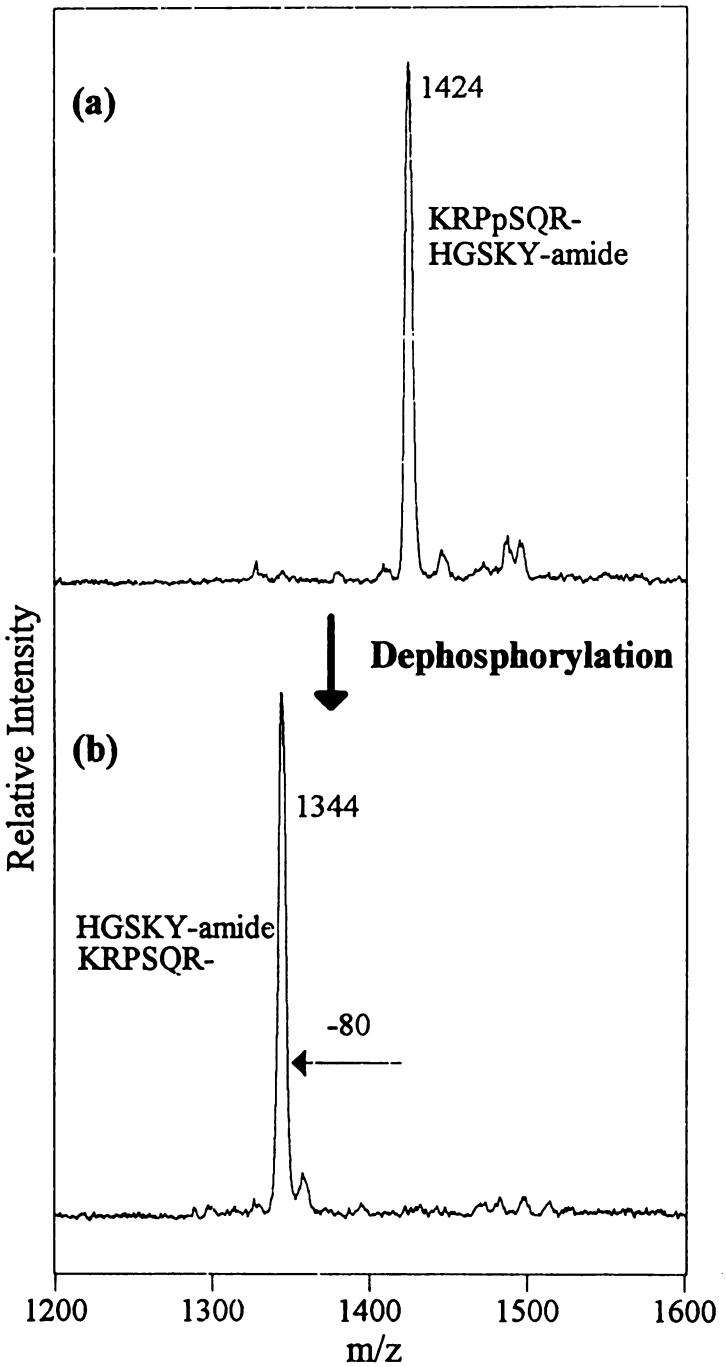


Figure 5.3 Enzymatic dephosphorylation reaction of immobilized phosphopeptide, KRPpSQRHGSKY-amide, on zetabind membrane for direct analysis by MALDI. The -80 Daltons shift in the mass spectral data is due to the loss of a phosphate moiety. 10 pmol of the peptide were used to perform this experiment.

This design is especially important when the peptide sample contains chemical substances which interfere with the subsequent dephosphorylation reaction and only a limited amount of sample is available. We have found that phosphopeptides purified from metal ion affinity columns may contain iron impurities which interfere with the dephosphorylation reaction. The problem was avoided by this sample immobilization and washing protocol.

Chapter 6 - Expression of Mouse Galectin-3 Using a Baculovirus/Insect Cell System

I. Introduction

In many animal cells, there is a protein called Galectin-3 (G3), present at very low levels. It is a lectin - a protein that interacts with carbohydrates. In some types of mice cells, it exists in two forms - free and monophosphorylated. The free form seems to be at elevated levels in cell nuclei when they begin to divide, so phosphorylation may be used to regulate the distribution of G3 inside and outside of the cell's nucleus.

Cell biologists would like to know where G3 is phosphorylated, but insufficient amounts can be isolated to make that determination. This is just one example of how small modifications, "post-translational modifications", are used in biological systems to regulate the activity of proteins/enzymes. Whether a protein is active or inactive may depend on whether it is phosphorylated, acetylated, or whether it undergoes any other type of derivatization. We became interested in questions involving phosphorylation of proteins, and found that very few were commercially available with which to work, in the context of MALDI-MS methods of analyses. Thus, we became involved in this topic, Galectin-3, to learn how one might make larger quantities for subsequent analysis.

Methods exist for using cells, such as *E. coli* cells, to generate ("express") large quantities of compounds such as G-3 for subsequent characterization. (Note, when this occurs, the product is referred to as rG-3, r = recombinant) However, bacteria probably would not phosphorylate the protein once it was made. So, we will use insect cells - since they contain the machinery that could phosphorylate such proteins once formed. We are

going to use a virus to convert an insect cell into a device that will generate G-3 and pG-3 (phosphorylated G-3) (Figure 6.1).

In the mouse cell, there is a portion of the DNA that is responsible for making G-3. A promoter assists in generating a messenger RNA (mRNA) from this portion of the DNA, and the mRNA is ultimately responsible for the generation of G-3. In the mouse cell, very little G-3 is required, thus the promoter is a "weak" promoter (Figure 6.1). Thus, the overall machinery is present for forming G-3, but only at low levels.

From the mRNA, we can create a piece of DNA that resembles that portion of the gene responsible for making G-3. We call that piece cDNA. We can add this cDNA, with a strong promoter, to a virus. It then becomes a "recombinant virus". This is then sent into an insect cell. The virus kills the cell, but we now have much of the cell machinery, a segment of cDNA, a strong promoter that will make copious amounts of the corresponding mRNA, which will then generate large amounts of G-3 (Figure 6.1). In addition to now having a machine that can generate G-3 at high levels, we also have machinery that can phosphorylate as well. In addition to ultimately generating G-3, the virus also reproduces itself, allowing for the "infection" of other cells. Thus, once the recombinant virus is engineered, a large number of insect cells that generate G-3 and pG-3 can be maintained.

The nomenclature for this area of biochemistry is very specific and very different from the nomenclature commonly used in chemistry. A description of the project, as it should be most appropriately and concisely described, follows.

Galectin-3 (M_r ~28,000) is a galactose-specific lectin found in the nucleus and cytoplasm of many animal cells [48]. Recent evidence indicates that it is a factor in the nuclear processing of pre-mRNA [49]. In mouse 3T3 fibroblast cells, Galectin-3 exists in two isoelectric forms: a pI (isoelectric point) 8.7 unmodified polypeptide and a pI 8.2

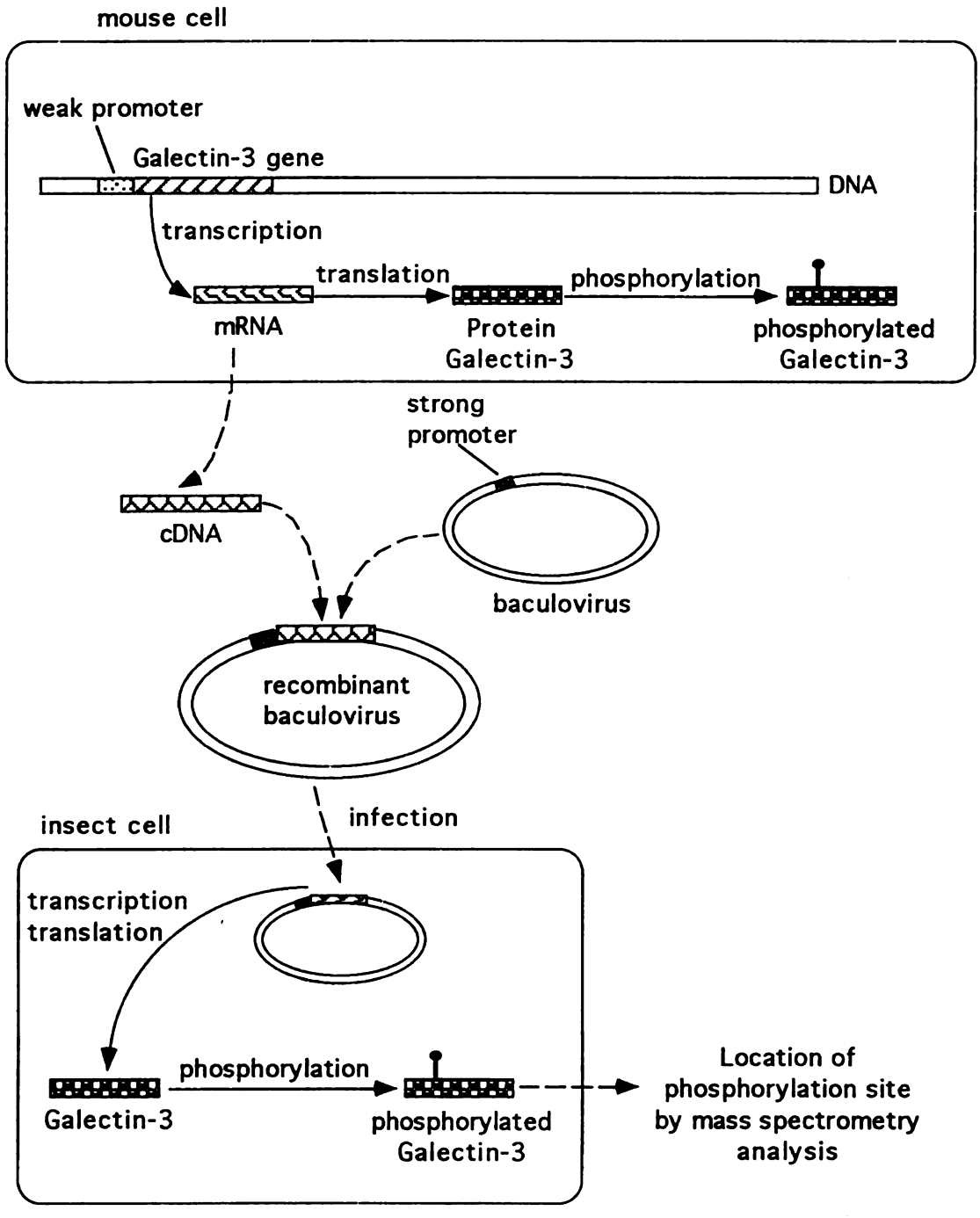


Figure 6.1 **Schematic diagram for the rationale and approach of producing large amounts of the phosphorylated form of Galectin-3 for the determination of the site of phosphorylation.** Galectin-3 is produced at low levels in mouse cells. The cloned cDNA of mouse Galectin-3 is engineered into the DNA of baculovirus; the recombinant baculovirus is used to infect its normal host (Sf21 insect cell). The virus takes over the protein synthesis and modification machinery of the host cell and produces viral proteins at high level, including phosphorylated mouse Galectin-3.

polypeptide modified by the addition of a single phosphate [50]. The pI 8.7 species is found exclusively in the nucleus and is the form that becomes highly elevated when cells are mitogenically stimulated into the proliferative state. Thus, it appears that phosphorylation of Galectin-3 may regulate its nuclear versus cytoplasmic distribution.

Because the level of Galectin-3 is very low in 3T3 cells (less than 0.01% of total cellular protein) [51], it would seem impractical to isolate a sufficient amount of the phosphorylated species to carry out chemical analysis for the identification of the site of phosphorylation. However, the cDNA for Galectin-3 has been cloned [52] and therefore, expression systems can be engineered to produce large amounts of the protein amenable for chemical studies [53]. Indeed, recombinant Galectin-3 has been produced in an *E. coli* expression system; the recombinant protein will be hereafter designated as rGalectin-3 (*E. coli*). However, bacteria generally do not phosphorylate proteins and thus rGalectin-3 (*E. coli*) has been shown to be homogeneously in the unphosphorylated (pI 8.7) form [50].

In the present study, we describe an alternative expression system that produces rGalectin-3 in animal cells. This system takes advantage of the fact that the cDNA of Galectin-3 can be engineered into an insect cell virus, baculovirus [54]. The recombinant virus is used to infect insect cells (Figure 6.1); the protein production and modification (including phosphorylation) machinery of the host cell is then used to produce viral proteins, including Galectin-3 of the recombinant virus.

II. Experimental

Construction of Baculovirus Expression Vector

Galectin-3 cDNA has been inserted into a plasmid; the product is pWJ31 [55]. Galectin-3 cDNA was isolated from pWJ31 by EcoRI digestion. The digest was electrophoresed on a 1.2 % agarose-TAE (40 mM Tris-acetate, 1 mM EDTA) gel and the

880 base pairs (bp) fragment was extracted using the Qiaex gel extraction kit (Qiagen Inc., Chatsworth, CA). The cDNA was subcloned into the baculovirus transfer vector pSynXIV VI$^+$ X3/2 (kindly provided by Dr. L. K. Miller, University of Georgia) using the EcoRI restriction site in the polylinker region. The ligated vector construct was transformed into XL-1 competent cells and selected on LB agar plates containing ampicillin (100 μg/mL). Individual colonies were picked, grown overnight in LB-ampicillin (100 μg/mL) and mini-prep DNA was isolated [56]. Positive clones were identified by EcoRI digestion of the DNA and subsequent agarose gel electrophoresis, yielding a 880 bp insert. The orientation of the insert relative to the ATG start codon in the transfer vector was determined by EcoRV digestion. EcoRV cleaves the cDNA insert 59 bp from the 5' end, and is also a single restriction site in the transfer vector. DNA preparations using QIAprep-spin plasmid kit (Qiagen Inc.) were made from large cultures of the transfer vector construct containing Galectin-3 in the correct orientation, pSynXIV VI$^+$ X3/2-Galectin-3, the transfer vector construct with Galectin-3 in the opposite orientation (antisense), pSynXIV VI$^+$ X3/2-AS Galectin-3, and the transfer vector alone, pSynXIV VI$^+$ X3/2, as a control.

Selection of Recombinant Virus

Spodoptera frugiperda (Sf21) insect cells were obtained from Dr. Suzanne Thiem (Michigan State University) and were grown as monolayer cultures at 27°C in TC-100 medium supplemented with 10 % fetal bovine serum (FBS). Sf21 cells (2 x 10^6) were seeded into three 60 mm dishes. After attachment (~ 1 hour), the culture medium was removed and 750 μL of Grace Insect medium was added to each plate. The cells were transfected using the calcium phosphate coprecipitation protocol [54]. The recombinant, vSynVI$^-$gal, an *Autographa california* multiply embedded nuclear polyhedrosis virus was obtained from Dr. L. K. Miller, University of Georgia. Plate A received 1 μg of viral DNA (vSynVI$^-$gal) and 2 μg of the transfer vector construct pSynXIV VI$^+$ X3/2-Galectin-3. Plate B received 1 μg of viral DNA (vSynVI$^-$gal) and

2 μg of the transfer vector construct pSynXIV VI$^+$ X3/2-AS Galectin-3. Plate C received 1 μg of viral DNA (vSynVI$^-$gal) and 2 μg of the transfer vector pSynXIV VI$^+$ X3/2. After 4 hours of incubation at 27° C, unadsorbed DNA was removed and 4 mL of culture medium was added. After 5 days, the supernatant from each plate was harvested and centrifuged 1000 x g for 5 minutes at 4°C to remove debris. Each supernatant was titered by plaque assay.

Plaque assays were performed by adding 0.5 mL of serial 10-fold dilutions of the supernatants to a dish containing monolayers of Sf21 cells (2 x 10^6 cells/dish). After one hour of incubation, unadsorbed virus was removed. An overlay of 4 mL of TC100 + 10 % FBS containing 0.5 % agarose (42°C) was slowly added to each plate. Plaques were visible by 3 - 4 days of incubation at 27°C. Plaques expressing the recombinant phenotype were picked, diluted in 1 mL of culture medium and isolated by three rounds of plaque purification. A stock of each recombinant virus was grown by infecting three 100 mm dishes seeded at a density of 5 x 10^6 cells/plate. The cells were infected at a multiplicity of infection (MOI) of 0.1. After 1 hour of incubation, 10 mL of culture medium were added. On day 5, the supernatants were removed, pooled, and centrifuged at 1000 x g for 5 min at 4°C. A plaque assay was performed to determine the titers of the stocks. All the stocks titered in the range of 10^7 plaque forming units (PFU)/mL. The stocks were designated vSynVI$^-$gal$^-$-Galectin-3, vSynVI$^-$gal$^-$-AS Galectin-3, and vSynVI$^-$gal$^-$.

Expression of Galectin-3

A 24 multi-well dish (2 cm^2/well) was seeded with 2 x 10^5 cells/well. After attachment, the culture medium was removed and 0.2 mL of diluted recombinant virus stocks (MOI ~ 2) was added to each well. Well 1A received 0.2 mL vSynVI$^-$gal$^-$-AS Galectin-3, well 2A received 0.2 mL of vSynVI$^-$gal$^-$-Galectin-3, and well 3A received 0.2 mL vSynVI$^-$gal$^-$. Following 1 hour of incubation at 27°C, 0.8 mL of culture medium was added. After 42 hours, the wells were washed twice with 1 mL of

PBS (1 mM Na_2HPO_4, 10.5 mM KH_2PO_4, 140 mM NaCl, 40 mM KCl; pH 6.2) and 25 μL of sodium dodecyl sulfate (SDS) sample buffer [57] was added to each well. The well was incubated at 37°C for 3 minutes to solubilize the cells and then the solubilized sample was carefully removed from the well. Each of the sample (10 μL) was subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) [57] (12.5 % acrylamide) and Western blotting.

Western Blotting

The proteins were transferred from SDS-PAGE gels onto Immobilon-P transfer membranes (Millipore Corporation, Bedford, MA) at 400 mA for 2 hours. The membranes were incubated at room temperature with blocking agent, 2 % gelatin in TBS (20 mM Tris, 0.5 M NaCl, 0.6 mM thimerosal; pH 7.5), for 1 hour and rinsed with T-TBS (0.05 % Tween-20 in TBS). The blocked membranes were incubated at room temperature with rabbit anti-Galectin-3 [53] diluted in 1% gelatin in TBS for two hours. After rinsing with T-TBS for 15 minutes three times, the membranes were incubated with goat anti-rabbit IgG-alkaline phosphatase (Boehringer Mannheim Biochemicals, Indianapolis, IN) diluted in 1% gelatin in TBS (1:5000) for two hours. Following three 15 minute washes with T-TBS, the membranes were developed with a substrate solution containing 15 mg p-NBT (p-nitro blue tetrazolium chloride) and 7.5 mg BCIP (5-bromo-4-chloro-3-indolyl phosphate p-toluidine salt) in 50 mL alkaline phosphatase buffer (0.1 M Tris, 0.1 M NaCl, 2 mM $MgCl_2$; pH 9.5). The reaction was stopped by rinsing with water.

Purification of rGalectin-3 (bv)

Four large plates (100 mm) were seeded with 5 x 10^6 Sf21 cells. After attachment, the medium (TC-100 + 10 % FBS) was removed and vSynVI$^-$gal$^-$-Galectin-3 was added at a MOI = 10 and diluted to 1 mL/plate. Following 1 hour of incubation at 27°C, the unadsorbed virus was removed and 10 mL of medium were added and incubated for 42 hours. The supernatant was removed and the cells were washed twice in cold PBS.

Cells were scraped into 2.5 mL of cold PBS and centrifuged for 5 minutes at 2000 rpm at 4°C. The supernatant was removed and the pelleted cells were resuspended in 2 mL of cold lysis buffer (50 mM Tris pH 7.5, 2 mM EDTA, 10 mM β-mercaptoethanol, 1 U/mL aprotinin, 1 μg/mL leupeptin, 1 mM phenylmethylsulfonyl fluoride (PMSF)) and incubated on ice for 20 minutes. The lysate was homogenized and sonicated three times for 15 seconds to break open the nuclei. The lysate was centrifuged 15 minutes at 2000 rpm at 4°C to pellet debris. The lysate was then transferred to a new tube, and the NaCl concentration was adjusted to 100 mM. The lysate was loaded onto a 0.5 mL column (0.5 cm x 2.5 cm) of α–lactose agarose equilibrated in binding buffer (50 mM Tris pH 7.5, 100 mM NaCl, 1 mM EDTA, 10 mM β-mercaptoethanol, 1 U/mL aprotinin, 1 μg/mL leupeptin, 1 mM PMSF) with a flow rate of 0.2 mL/minute. The flow through was reloaded onto the column 2 more times. The column was then washed with 16 volumes of binding buffer. The column was eluted with lactose elution buffer (50 mM Tris pH 7.5, 0.4 M lactose, 10 mM β-mercaptoethanol, 1 U/mL aprotinin, 1 μg/mL leupeptin, 1 mM PMSF), collecting 0.5 mL per fraction. The column was extensively washed in Tris buffer (50 mM Tris pH 7.5, 0.2 % sodium azide) and stopped. Samples from each fraction were subjected to SDS-PAGE, silver staining, and Western blotting.

Two-dimensional Gel Electrophoresis

The first dimension was established by isoelectric focusing (IEF) and the second dimension by SDS-PAGE. These procedures have been reported by O'Farrell [58]. The apparatus used was an Investigator™ 2-D Electrophoresis System (Millipore Corporation, Bedford, MA) [59]. The IEF tube gels (1mm x 18 cm) were prepared using pH 3-10 ampholines, prefocused (1,500 Volt, 2.5 hours), and run for 18 hours using 18,000 Volt-hours (maximum voltage: 2000 V, maximum current: 110 mA). 750 ng of rGalectin-3 (*E. coli*) and 750 ng rGalectin-3 (bv) were loaded onto two separate IEF gels. The pH gradient was measured by cutting a duplicate (dummy) gel into 1 cm slices and the

pH in ea

second

protocol

Reverse

MALD

in Appa

III. Re

Constru

cDNA

pWJ31

because

[54]. In

virus. T

fragme

transcri

The rec

S21 ce

transfe

recomb

exchan

recomb

their re

recomb

pH in each slice was measured after equilibrating with 0.7 mL water for an hour. The second dimension SDS-PAGE contained 12.5 % acrylamide. A modified Morrissey protocol [59,60] was used to silver stain the two- dimensional gels.

Reverse Phase High Performance Liquid Chromatography (HPLC) Fractionation and MALDI-TOF-MS analysis

HPLC, mass spectrometry, and experimental conditions were the same as described in Appendix IV, except that sinapinic acid was used as the matrix for MALDI.

III. Results and Discussion

Construction of Recombinant Baculovirus

The construction of recombinant baculovirus is illustrated in Figure 6.2. A 880 bp cDNA fragment containing the coding region of mouse Galectin-3 was isolated from pWJ31 [55] by digestion with EcoRI. The transfer vector pSynXIV VI^+ X3/2 was used, because it contains the necessary initiating ATG codon followed by a unique EcoRI site [54]. It also contains a polyhedrin gene which will be used for the selection of recombinant virus. The recombinant transfer vector was constructed by inserting the Galectin-3 cDNA fragment into the EcoRI site of pSynXIV VI^+ X3/2 in the correct orientation, so that the transcription of Galectin-3 gene is under the control of a strong hybrid promoter P_{synXIV}. The recombinant transfer vector (pSynXIV VI^+ X3/2-Galectin-3) was cotransfected into Sf21 cells with the parent viral DNA (vSynVI^-gal) using a calcium phosphate-mediated transfection procedure [54]. Although it occurs with low frequency (< 5%), homologous recombination can take place, in which segments of the recombinant transfer vector exchanges DNA with the parent viral DNA. In the present case, the segment of the recombinant transfer vector bounded by the Galectin-3 and polyhedrin gene (containing their respective promoters) replaces a LacZ gene in the parent viral DNA, resulting in a recombinant virus (Figure 6.2).

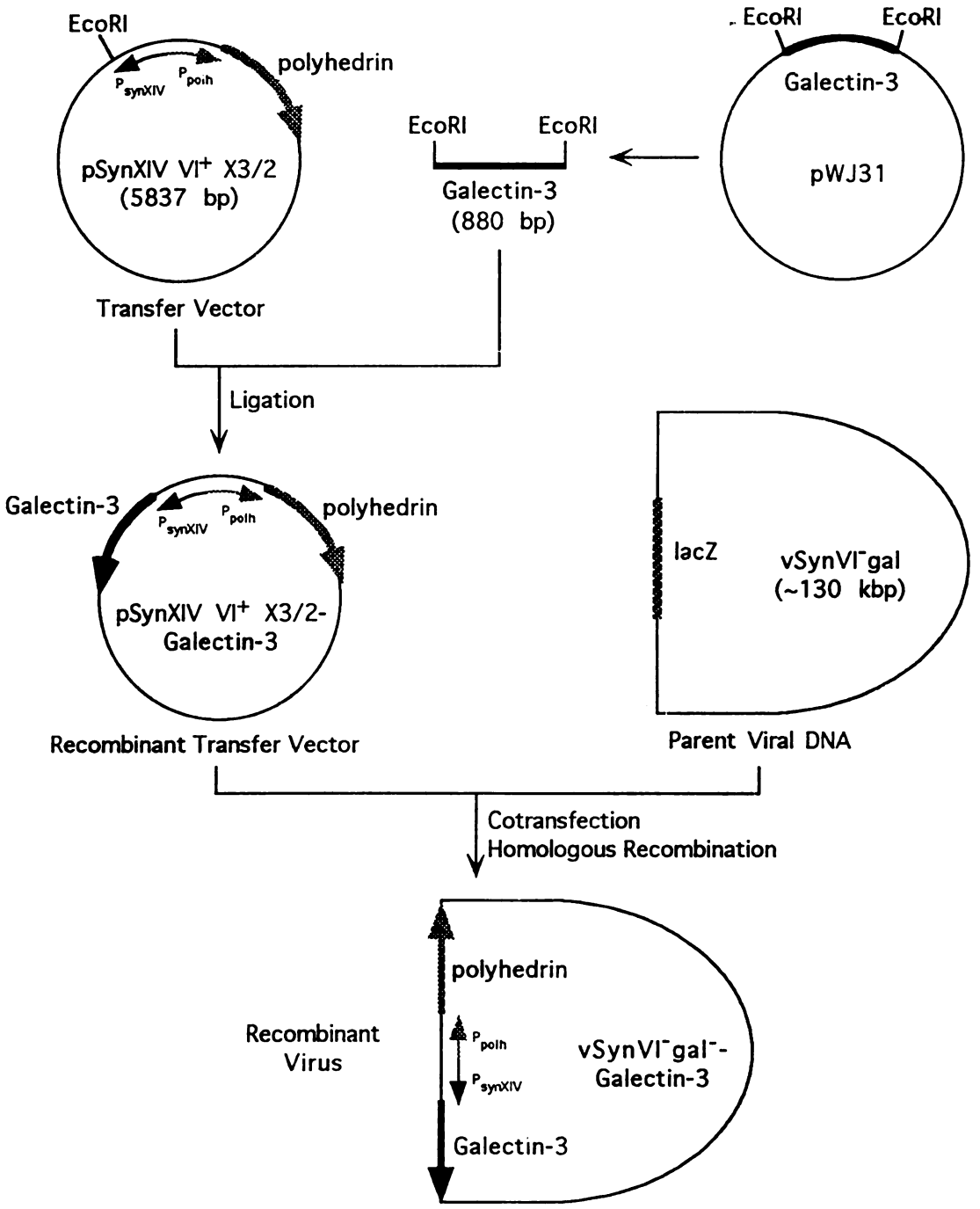


Figure 6.2 **Schematic diagram of the construction of the recombinant baculovirus vSynVI⁻gal⁻-Galectin-3.** The cDNA for Galectin-3 was isolated from plasmid pWJ31 and subcloned into the EcoRI sites of the transfer vector pSynXIV VI+ X3/2 immediately downstream from the ATG initiating codon. After cotransfection of the parent baculovirus DNA and the transfer vector construct DNA into the Sf21 cells, the recombinant baculovirus was identified and selected by the recombinant phenotypes.

occlu

plaqu

infec

a chr

selec

stock

Expre

42 ho

herea

and V

6.3, r

There

(*E. co*

for m

immu

polyp

(b) a

endog

not be

ATG

onto a

linker

termin

The resulting recombinant viruses were selected on the basis of both the positive occlusion phenotype and the absence of blue plaque phenotype. The former is a distinct plaque morphology, which is formed by the existence of polyhedrin occlusion bodies in the infected cells. The latter results in white plaques in the blue background in the presence of a chromogenic indicator, due to the absence of LacZ gene in the recombinant virus. The selected viruses were subjected to three rounds of plaque purification to produce a pure stock of recombinant viruses (vSynVI$^-$gal$^-$-Galectin-3).

Expression of Recombinant Galectin-3 in vSynVI$^-$gal$^-$-Galectin-3-infected Insect Cells

Sf21 cells were infected with recombinant viruses with a low MOI (~ 2) for 42 hours. The expression of Galectin-3 coded by the recombinant viruses, which is hereafter designated as rGalectin-3 (bv), was examined by one-dimensional SDS-PAGE and Western blotting analyses of extracts derived from these cells. As shown in Figure 6.3, rGalectin-3 (bv) was detected by immunoblotting with rabbit anti-Galectin-3 (lane 4). There was a predominant band with an electrophoretic mobility identical to rGalectin-3 (*E. coli*) which was electrophoresed in parallel (Figure 6.3, lanes 1-3), both as a standard for molecular weight and as a standard for quantification (see below). Three minor bands immunoreactive with the antibody were also observed in lane 4 (Figure 6.3): (a) a ~ 14 kD polypeptide, most probably representing a proteolysis fragment of rGalectin-3 (bv); (b) a ~ 38 kD polypeptide, which is suspected to be the insect homolog of Galectin-3, endogenous to the Sf21 cells (see below); and (c) ~ 70 kD polypeptide whose identity has not been defined.

The first amino acid of the protein product is methionine, provided by the initiating ATG codon on the transfer vector pSynXIV VI$^+$ X3/2. Two extra amino acids are added onto the amino terminus of the polypeptide coded by Galectin-3 cDNA, as a result of the linker (EcoRI cleavage site) which is next to the ATG initiating codon, Thus, the amino-terminal end of rGalectin-3 (bv) has the following sequence: Met-Glu-Phe-Arg-Asp-Ser-,

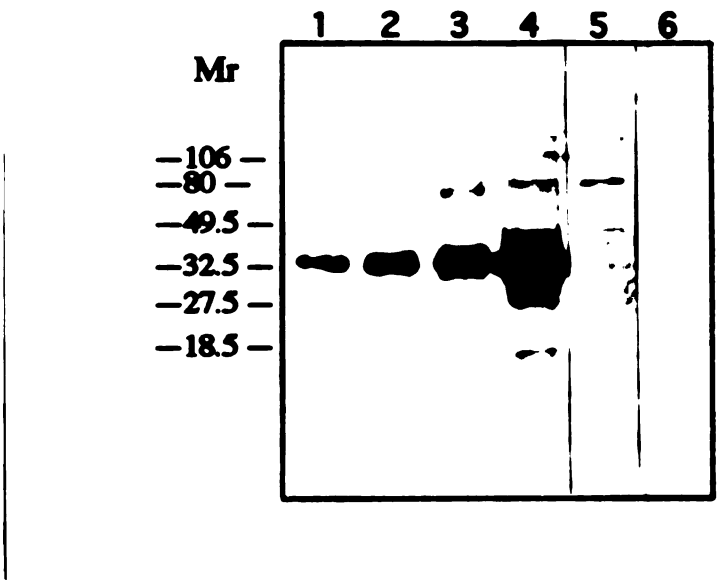


Figure 6.3 **Expression of Galectin-3 in recombinant virus-infected insect cells as assayed by one-dimensional SDS-PAGE analysis.** Sf21 cells were infected with recombinant virus. After 42 hours of incubation, the cells were solubilized in sample buffer, and subjected to SDS-PAGE. The proteins were revealed by immunoblotting with rabbit anti-Galectin-3. Lane 1-3 contain, respectively, 3 ng, 15 ng, and 30 ng of purified rGalectin-3 (*E. coli*), which serve as a reference. Lane 4: 10 µl of total extracts of cells infected with recombinant viruses, vSynVI$^-$gal$^-$-Galectin-3, containing the cDNA of Galectin-3 in the correct orientation. Lane 5: 10 µl of total extracts of cells infected with control recombinant viruses, vSynVI$^-$gal$^-$-AS Galectin-3, containing the cDNA of Galectin-3 in the antisense orientation. Lane 6: 10 µl of total extracts of cells infected with control recombinant viruses, vSynVI$^-$gal$^-$, containing the recombinant region of the transfer vector only. The numbers on the left indicate the positions of migration of molecular weight standards.

with the fourth residue (Arg) being the first amino acid from Galectin-3 cDNA. According to the published sequence of the cDNA and the three additional amino acids added at the amino terminus, the calculated molecular weight of rGalectin-3 (bv) is 27,814. On SDS-PAGE, however, the mobility of rGalectin-3 (bv) corresponds to that of a polypeptide of M_r ~ 35,000 (Figure 6.3, lane 4). This anomaly of electrophoretic mobility was also observed with endogenous Galectin-3 of mouse 3T3 fibroblasts [51,52].

Since the cDNAs for the coding regions used in the expression of Galectin-3 in *E. coli* and in baculovirus were essentially the same, the reactivity of the antibody with rGalectin-3 (*E. coli*) and rGalectin-3 (bv) was expected to be comparable. This formed the basis for estimating the amount of rGalectin-3 (bv) expressed in the Sf21 cells. Different amounts of purified rGalectin-3 (*E. coli*) were electrophoresed and immunoblotted (Figure 6.3, lane 1-3). The intensities of the Galectin-3 band were found to be directly proportional to the amount of rGalectin-3 (*E. coli*). Assuming that this proportionality held, even at large amounts of protein electrophoresed, the amount of rGalectin-3 (bv) in the lane 4 (Figure 6.3) was estimated to be ~ 100 ng. The amount of extract analyzed in lane 4 represented 40 % of the total extract isolated from 2 x 10^5 infected Sf21 cells. Therefore, 1.25 μg of rGalectin-3 (bv) are produced by 1 x 10^6 cells under these conditions of infection (low MOI).

Sf21 cells were also infected with recombinant viruses containing the cDNA of Galectin-3 in the antisense orientation and with viruses containing the transfer vector alone. Extracts derived from these cells yielded no detectable mouse Galectin-3 (Figure 6.3, lanes 5 and 6). In these analyses, however, a faint band was often observed at M_r ~ 38,000 (see for example, Figure 6.3, lanes 4 and 5). This band was observed only in Sf21 cell extracts, irrespective of whether the cells were infected with viruses. On the basis of its carbohydrate-binding activity (P. G. Voss and P. C. Liao, unpublished observations)

and immunoreactivity, this polypeptide is suspected to correspond to the insect homolog of Galectin-3 endogenous to Sf21 cells.

Carbohydrate-binding Activity of rGalectin-3 (bv)

Extracts of Sf21 cells, infected with recombinant viruses with an MOI of 10 for 42 hours, were fractionated by affinity chromatography on a column of lactose-agarose beads. Material bound to the column was subsequently eluted with lactose. When subjected to SDS-PAGE and silver staining, this fraction yielded a M_r ~ 35,000 polypeptide as the predominant component (Figure 6.4a, lane 3). This polypeptide could be immunoblotted with anti-Galectin-3 (Figure 6.4b, lane 3). It exhibited the same electrophoretic mobility as purified rGalectin-3 (*E. coli*) (Figure 6.4a and 6.4b, lane 1). On the basis of these observations, we conclude that rGalectin-3 (bv) is bound to the carbohydrate, lactose. This binding allowed the purification of rGalectin-3 (bv) from the myriad of polypeptides present in the unfractionated extract (Figure 6.4a, lane 2).

In addition to the predominant band at M_r ~ 35,000, two other bands were also observed in the ~ 70 kD region of the silver-stained gel (Figure 6.4a, lane 3). These two bands failed to react with anti-Galectin-3 (Figure 6.4b, lane 3). However, these two bands were also observed in the purified rGalectin-3 (*E. coli*) sample (Figure 6.4a, lane 1). Although the identities of these two bands remain to be established, we surmise that they represent "finger proteins", keratin components of the skin on the finger which have contaminated the SDS-PAGE buffers.

Using an approach similar to that previously described for antibody blotting, the amount of rGalectin-3 (bv) purified from this system was estimated to be ~ 3.5 μg per 10^6 cells.

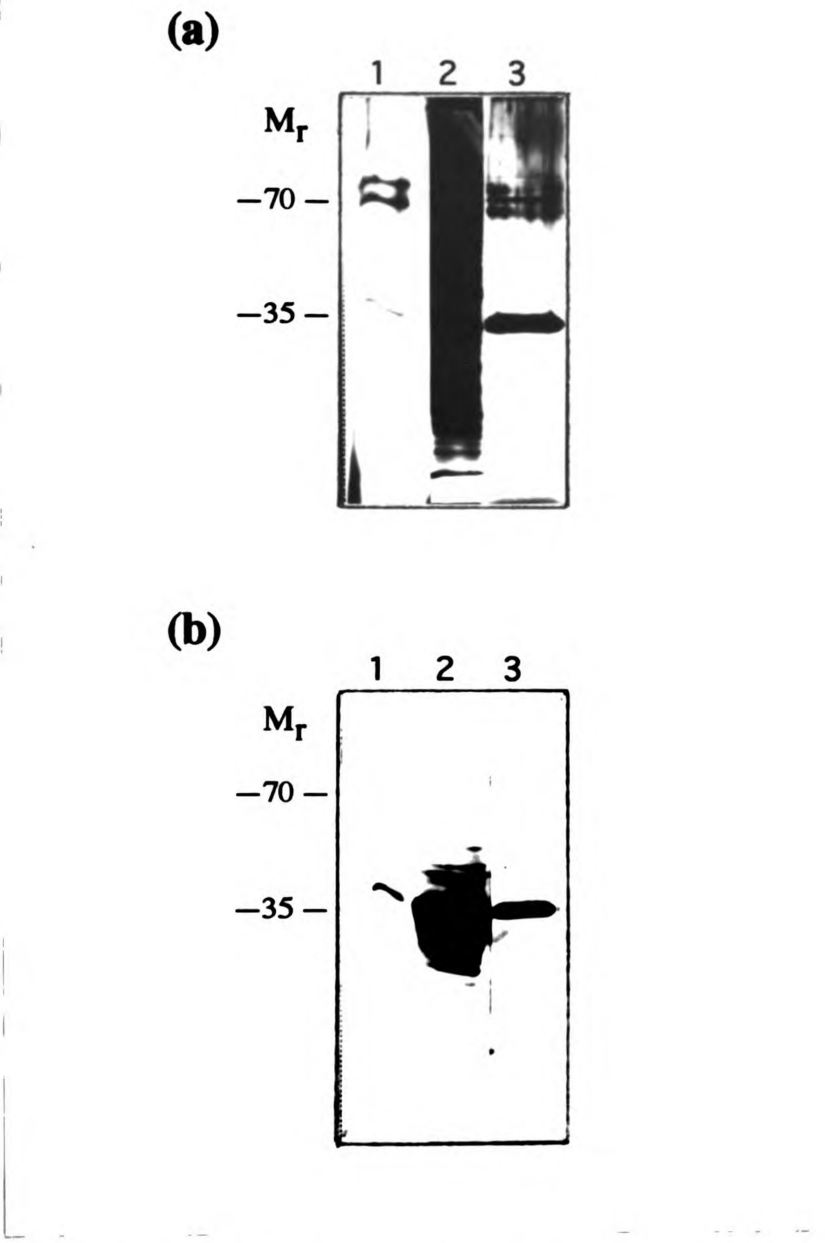


Figure 6.4 **Determination of carbohydrate-binding activity of rGalectin-3 (bv) by SDS-PAGE analysis of fractions before and after affinity chromatography.** Total extracts of Sf21 cells infected with recombinant virus, vSynVI$^-$gal$^-$-Galectin-3, were loaded onto a lactose-agarose column. Bound material was eluted with 0.4 M lactose. Samples were analyzed by SDS-PAGE, which were either (a) silver stained, or (b) immunoblotted with rabbit anti-Galectin-3. Lane 1: 50 ng of purified rGalectin-3 (*E. coli*). Lane 2: 20 µl of lysate derived from recombinant virus-infected cells prior to affinity column purification. Lane 3: 20 µl of lactose-eluted fraction from the affinity column. The position of migration of polypeptide of M_r ~ 35,000 and ~ 70,000, corresponding to the gel shown in Figure 6.3, are highlighted on the left.

Phosphorylation of rGalectin-3 (bv) in Insect Cells

When extracts of mouse 3T3 cells were subjected to two-dimensional gel electrophoresis and immunoblotting by anti-Galectin-3, two spots were observed, corresponding to pI 8.7 and 8.2. The pI 8.2 species represents a posttranslational product, by addition of a single phosphate group to an unmodified pI 8.7 polypeptide [50]. The rGalectin-3 (bv) purified by the affinity chromatography (Figure 6.4, lane 3) and rGalectin-3 (*E. coli*) [53] were analyzed by two-dimensional gel electrophoresis: IEF followed by SDS-PAGE and silver staining (Figure 6.5). In Figure 6.5a, rGalectin-3 (*E. coli*) yielded a single spot corresponding to a pI ~ 8.6. In contrast, two spots are produced by rGalectin-3 (bv), corresponding to a pI ~ 7.9 and pI ~ 8.6 (Figure 6.5b). The pH range of this particular set of IEF gels did not extend beyond pH 8.7; therefore, the pI values assigned for the observed spots in Figure 6.5 are only estimates. We believe that the pI ~ 8.6 spot seen in both Figure 6.5 a and 6.5b corresponds to the unmodified Galectin-3 polypeptide and that the pI ~ 7.9 spot observed in the rGalectin-3 (bv) sample corresponds to the phosphorylated form of the polypeptide. This indicates that the recombinant rGalectin-3 (bv) is partially phosphorylated in the host cells. The ratio of non-phosphorylated to phosphorylated forms of rGalectin-3 (bv) in infected Sf21 cells was estimated to be ~ 10:1.

MALDI-TOF-MS analysis of rGalectin-3 (bv)

The rGalectin-3 (bv) purified from affinity chromatography (Figure 6.4, lane 3) was desalted and fractionated by the reverse phase HPLC to obtain a pure rGalectin-3 (bv) sample for MALDI analysis. Figure 6.6a shows the MALDI-TOF mass spectral data obtained by using 1 pmol of rGalectin-3 (bv) (estimated by UV absorption at 214 nm) and 2 pmol of insulin as an internal calibrant. The average molecular mass of rGalectin-3 (bv) is expected to be 27,814 and is measured as 27,806 Daltons using the centroid of the peak. In Figure 6.6b, the mass spectral region near the peak corresponding to rGalectin-3 (bv)

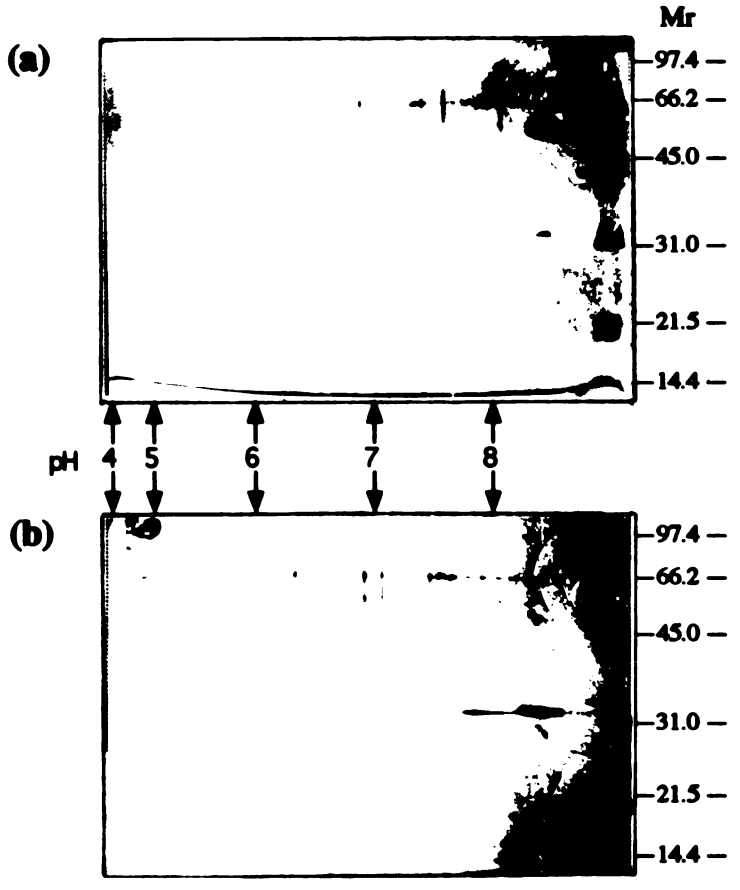


Figure 6.5 **Isoelectric points of rGalectin-3 (*E. coli*) and rGalectin-3 (bv) as determined by two-dimensional gel electrophoretic analysis.** (a) 750 ng of rGalectin-3 (E. coli); (b) 750 ng of rGalectin-3 (bv). The pH range of the isoelectric focusing gel in the first dimension is shown on the horizontal axis. The molecular weight range of the SDS-PAGE in the second dimension is shown in the vertical axis.

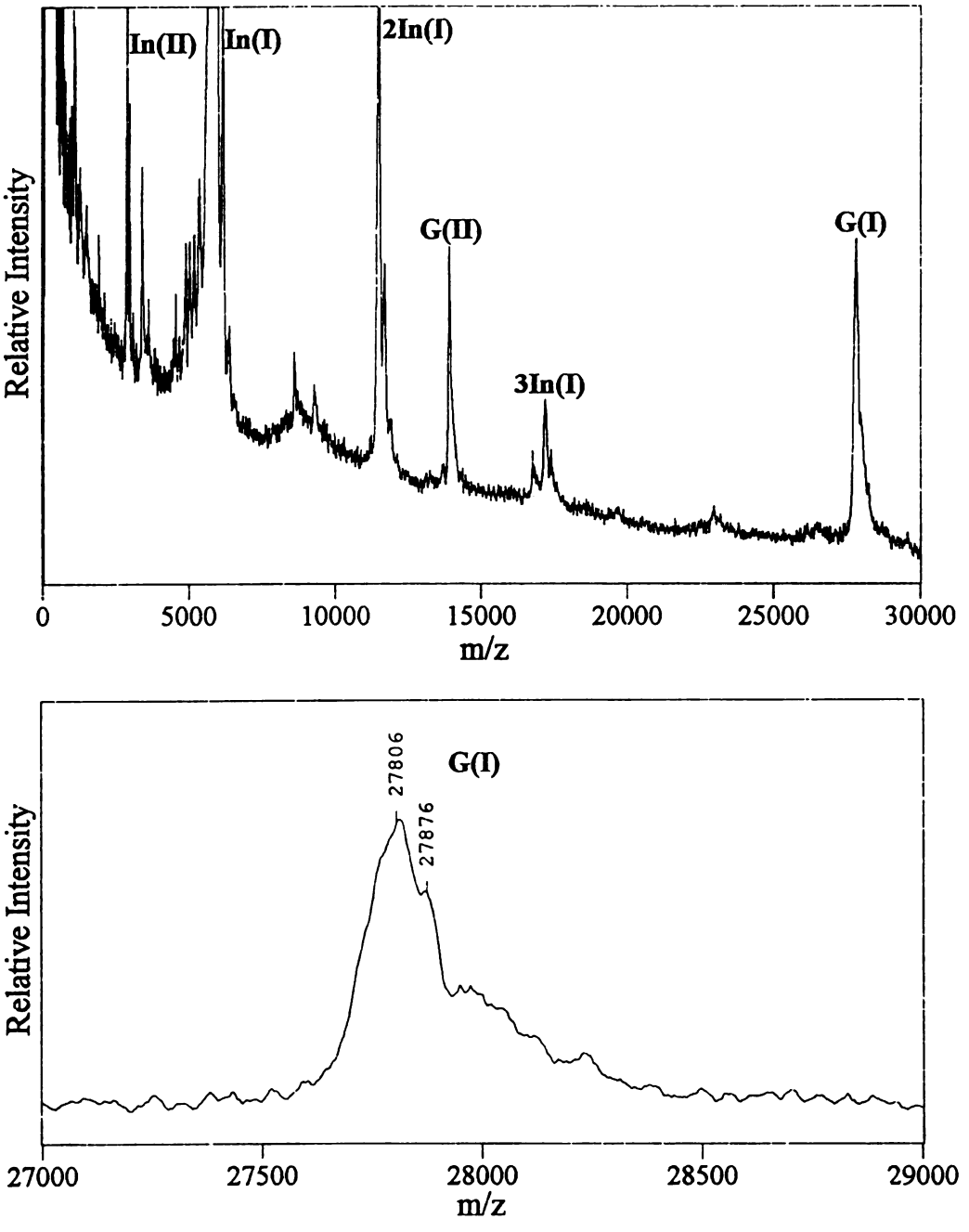


Figure 6.6 **MALDI-TOF mass spectral analysis of rGalectin-3 (bv).** (a) The data were obtained by from 1 pmol of rGalectin-3 (bv) and 2 pmol of insulin as an internal calibrant. (b) The mass spectral region near the peak corresponding to rGalectin-3 (bv) analyte is shown. Abbreviations: In, insulin; G, rGalectin-3 (bv); (I), singly-charged; (II), doubly-charged; 2In, insulin dimer.

analyte is shown. The satellite peak at the high mass end results from adduct formation with a dehydrated sinapinic acid matrix molecule (MW of sinapinic acid: 224). The high mass shoulder of the peak may indicate that the rGalectin-3 (bv) exhibits certain heterogeneity which may be constituted by the phosphorylated form of rGalectin-3 (bv) (expected mass shift due to the addition of a phosphate moiety: 80 Daltons). However, this requires further investigation.

IV. Conclusions

Galectin-3 has been successfully produced using a baculovirus/insect cell expression system. The mass of the mature protein, rGalectin-3 (bv), is confirmed by MALDI-TOF-MS analysis. rGalectin-3 (bv) also exhibits carbohydrate-binding activity and is shown to be phosphorylated in the insect cells as indicated by two-dimensional electrophoresis analysis. The next step of this study is to locate the phosphorylation site(s) of phosphorylated rGalectin-3 (bv).

In Chapter 5, the development of the method to locate phosphorylation sites in a phosphoprotein using MALDI-TOF-MS has been described. The sample consumption of this method is currently limited by enzymatic digestion and chromatographic separation of the digestion mixture. It is estimated that about 3 μg, or 100 pmol, of phosphorylated rGalectin-3 (bv) will be required to carry out a successful analysis. Since the method does not require a sample of pure phosphorylated form, the rGalectin-3 (bv) purified by affinity chromatography can be used and subjected to enzymatic digestion. In Figure 6.5b, it is shown that ~ 10 % of the rGalectin-3 (bv) is phosphorylated in the insect cells. That is, at least 30 μg of the purified rGalectin-3 (bv) are required, which can be easily produced from a large stock of recombinant virus-infected insect cells.

Appendix I

JOURNAL OF MASS SPECTROMETRY. VOL. 30. (1995)

Ionization Processes in Matrix-assisted Laser Desorption/Ionization Mass Spectrometry: Matrix-dependent Formation of $[M + H]^+$ vs $[M + Na]^+$ Ions of Small Peptides and Some Mechanistic Comments

Pao-Chi Liao and John Allison†
Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, USA

The relative signal intensities of $[M + H]^+$ vs $[M + Na]^+$ ions of some small peptides were found to be highly matrix-dependent in matrix-assisted laser desorption/ionization (MALDI) experiments when sinapinic acid, α-cyano-4-hydroxycinnamic acid and 2,5-dihydroxybenzoic acid were used as matrices. Presumably, this observation results from the competition of two different mechanisms for the ionization steps. Possible mechanisms for the formations of $[M + H]^+$ and $[M + Na]^+$ ions are proposed and discussed in the pursuit of defining the ionization step in MALDI. Several experiments were designed and performed for the further refinement of the mechanism. The results suggest that proton transfer from a matrix molecule to an analyte plays an important role in the ionization step. The transferring proton may be derived from photoionized or electronically excited matrix molecules. In contrast, some data are most consistent with a gas-phase mechanism for $[M + Na]^+$ ions.

INTRODUCTION

Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS), developed by Karas *et al.*,[1] is a sensitive technique for the analysis of a variety of large biomolecules[2–9] and synthetic polymers[10,11] with molecular masses exceeding 10^5. MALDI has one feature that links it to other desorption/ionization techniques such as fast atom bombardment (FAB) and secondary ion mass spectrometry (SIMS), namely that its development and commercial availability preceded our understanding of the chemical and physical processes that are involved. To generate ions in MALDI, analytes must be desorbed and, at some point, ionized. Models to describe the desorption of large, intact molecules from an energized matrix have been pursued,[12] and appear to describe accurately the desorption step. The chemical step, leading to ionization, is still a matter of considerable debate and research. Our initial interest was to attempt to understand why some matrices seem to generate $[M + H]^+$ ions for peptide analytes, M, while other matrices yield predominantly $[M + Na]^+$ ions for the same analytes. In this context, possible mechanisms of the ionization step in MALDI were evaluated. We concentrate our efforts here on positive-ion formation, using UV laser excitation.

* Paper prepared for publication in *Organic Mass Spectrometry*.
† Author to whom correspondence should be addressed.

EXPERIMENTAL

All MALDI results were obtained on a Vestec VT2000 linear time-of-flight (TOF) mass spectrometer (Vestec, Houston, TX, USA) equipped with a Model VSL-337ND nitrogen laser (Laser Science, Newton, MA, USA) (337 nm, 3 ns pulse, 250 μJ per pulse). Laser light was attenuated using a Model 935-5-OPT computer-controlled variable attenuator (Newport, Fountain Valley, CA, USA). The accelerating voltage in the ion source was 30 kV. Data were acquired with a transient recorder with 5 ns resolution. For the matrices sinapinic acid and α-cyano-4-hydroxycinnamic acid, water–acetonitrile (2:1, v/v) was used as the solvent system to prepare the matrix solution. When 2,5-dihydroxybenzoic acid was the matrix, water was used as the solvent. Peptides and other organic analytes were dissolved using an aqueous 0.1% trifluoroacetic acid–acetonitrile (2:1 or 1:1, v/v) solvent. To prepare the sample, 1 μl of the analyte solution was added to 1 μl of the matrix solution and applied to a flat stainless-steel probe tip. The mixture was then allowed to air dry before introduction into the ion source. All spectra were obtained under similar conditions in that the laser power used was just above the threshold for ion formation. Each spectrum shown represents the sum of data from 64 laser shots to improve the signal-to-noise ratio. Time-to-mass conversion was achieved by either external or internal calibration using peaks for Na^+ and K^+, most abundant matrix peaks and insulin peaks ($[M + H]^+$ and $[M + 2H]^{2+}$).

CCC 1076–5174/95

Received 12 August 1994
Revised manuscript received/1994
Accepted/1994

All peptides samples (hexaglycine, hexaalanine, pentaphenylalanine, hexatyrosine and H-VGVAPG-OH) were purchased from the Sigma Chemical (St Louis, MO, USA). All matrices (sinapinic acid, α-cyano-4-hydroxycinnamic acid and 2,5-dihydroxybenzoic acid) and organic analytes (4-cyanobenzaldehyde, thymidine, phenazine, acridine, 1,7-diaminoheptane and 1,8-bis(dimethylaminolnaphthalene) were purchased from Aldrich Chemical (Milwaukee, WI, USA). The compounds were used as purchased without further purification unless specified. Cation-exchange resin (AG 50W-X8, mesh size 100–200) was purchased from Bio-Rad Laboratories (Melville, NY, USA).

To desalt the analyte or matrix, solutions of hexatyrosine and α-cyano-4-hydroxycinnamic acid were specifically desalted for some experiments, by exposing them to hydrogen-bound cation-exchange resin for 24 h. The supernatant liquid was collected after centrifugation to remove the resin. The concentration of α-cyano-4-hydroxycinnamic acid purified in this way was determined by UV spectrophotometry at 326 nm, and the concentration of desalted hexatyrosine was determined by reversed-phase HPLC using UV detection at 214 nm. The titration curve that is presented was constructed by plotting log {(peak intensity for $[M + Na]^+$)/(peak intensity for $[M + H]^+$)} vs. log (number of picomoles of NaCl in sample on probe) to study the formation of sodiated analytes. In this experiment, the components on the probe were determined to 6000 pmol of α-cyano-4-hydroxycinnamic acid and 40 pmol of hexatyrosine.

The methyl ester of pentaphenylalanine was prepared according to the method described by Knapp.[13] A solution of 2 M methanolic HCl was prepared and then added to 1 nmol of dry pentaphenylalanine. The excess reagent was removed by evaporation after a reaction time of 100 min at room temperature.

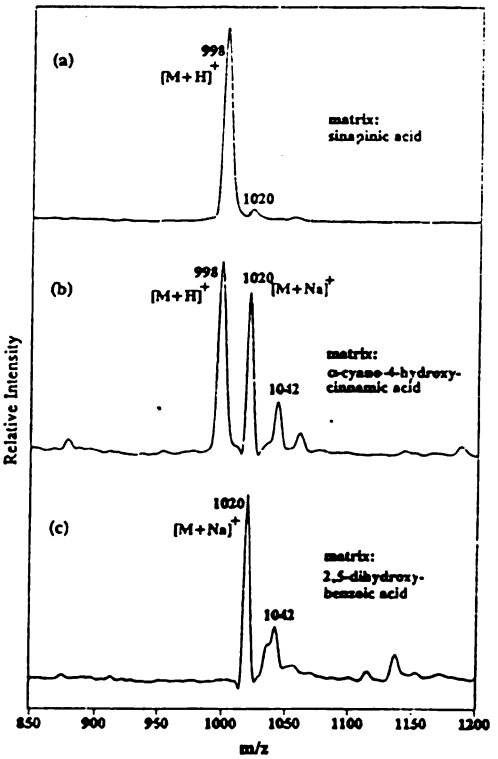


Figure 1. Molecular ion region of three MALDI mass spectra. In each case, the analyte is hexatyrosine, and excess NaCl has been added to the matrix. Peaks observed represent the protonated molecule at *m/z* 998, the $[M + Na]^+$ species at *m/z* 1020 and $[M - H + 2Na]^+$ at *m/z* 1042. In (b) and (c), the low-mass shoulder on the *m/z* 1042 peak is due to $(M + K)^+$, *m/z* 1036.

RESULTS AND DISCUSSION

We were intrigued by the fact that MALDI-MS samples using some matrices lead to the formation of alkali metal ion adducts of peptides whereas other matrices favor the formation of protonated molecules, and began to pursue some experiments designed to provide insights into these processes. Of course, one explanation may be that some matrices contain NaCl whereas others do not. The simplest experiment involved adding NaCl to a variety of matrices. Even with excess NaCl in the sample, some matrices yield protonated molecules and others sodiated molecules. An example is shown in the three MALDI mass spectra of hexatyrosine that are shown in Fig. 1. In Fig. 1(a), with sinapinic acid as the matrix, the pseudomolecular ion is the $[M + H]^+$ ion. In Fig. 1(b), with α-cyano-4-hydroxycinnamic acid as the matrix, both an $[M + H]^+$ and an $[M + Na]^+$ peak are present. In Fig. 1(c), with 2,5dihydroxybenzoic acid as the matrix, only an $[M + Na]^+$ peak is present. Excess sodium chloride had been added to each sample to give a sodium chloride (added):matrix:analyte molar ratio of 125:50:1. Although this is an extreme amount of NaCl to add, the three spectra are similar to those obtained using the three matrices as supplied from the manufacturer (*without* adding NaCl). Thus, we find it intriguing that, even with such a large amount of NaCl present, sinapinic acid only yields protonated analyte molecules.

Before designing experiments that might provide insights into the formation of $[M + H]^+$ and $[M + Na]^+$ ions in MALDI, possible mechanisms for their formation were considered, and are presented here. These are followed by some experimental observations, and their implications concerning the chemical aspects of the MALDI technique. We shall focus here on the chemical aspects leading to positive-ion formation.

Possible mechanisms for the formation of $[M + H]^+$ ions in MALDI

For this mechanistic discussion, we shall designate the matrix molecules as mH (specifically indicating the presence of an H atom in the molecule) and the analyte molecules as M. This designation leads to the simplest description of variants of each that may be involved in the chemistry of the MALDI experiment.

Mechanism 1: excited-state acid–base chemistry. In this mechanism the electronically excited matrix molecule, mH^*, formed in reaction (1a), acts as an acid, transferring a proton to surrounding matrix and/or peptide molecules. This could occur either prior to or following desorption, but presumably such processes are most likely immediately following excitation, when the number of mH^* present is highest, in the condensed phase. In this mechanism, quantities which may influence analyte ionization may include the proton affinity (PA) of the analyte and the heterolytic bond dissociation energy, $BDE(m^-—H^+)$ of the matrix.

$$mH + h\nu \rightarrow mH^* \quad (1a)$$

$$mH^* + M \rightarrow m^- + [M + H]^+ \quad (1b)$$

$$mH^* + mH \rightarrow m^- + [mH + H]^+ \quad (1c)$$

Mechanism 2: proton transfer following matrix photoionization Following matrix photoionization in reaction (2a), the ionized matrix molecule protonates the analyte (reaction (2b)) and/or another matrix molecule (reaction (2c)).

$$mH + n h\nu \rightarrow mH^{+\cdot} + e^- \quad (2a)$$

$$mH^{+\cdot} + M \rightarrow m^{\cdot} + [M + H]^+ \quad (2b)$$

$$mH^{+\cdot} + mH \rightarrow m^{\cdot} + [mH + H]^+ \quad (2c)$$

Apart from direct proton transfer following photoionization, there is an alternative mechanism which resembles chemical ionization (CI), mechanism 2'.

Mechanism 2': proton transfer sequence following matrix photoionization.

$$mH + n h\nu \rightarrow mH^{+\cdot} + e^- \quad (2'a)$$

$$mH^{+\cdot} + mH \rightarrow m^{\cdot} + [mH + H]^+ \quad (2'b)$$

$$[mH + H]^+ + M \rightarrow mH + [M + H]^+ \quad (2'c)$$

In the sequence of steps in reactions (2'b) and (2'c), the ionized matrix molecule reacts with another matrix molecule to form a protonated matrix molecule, which is the species responsible for protonation of the analyte. This mechanism is considered since the molar matrix-to-analyte ratios in MALDI are large, as are reagent gas-to-analyte ratios in CI. It is also considered because all of the ionic species involved in mechanism 2' are detected.

In these mechanisms (2 and 2'), quantities which may influence successful analyte ionization may include the proton affinity of the analyte, the ionization energy (IE) of the matrix and $BDE(m^{\cdot}—H^+)$.

Mechanism 3: H-atom transfer following analyte photoionization. In this model, the energy deposited into surrounding excited matrix molecules is used to ionize a peptide molecule, which subsequently extracts a hydrogen atom from a matrix molecule.

$$mH + h\nu \rightarrow mH^* \quad (3a)$$

$$n mH^* + M \rightarrow n mH + M^{+\cdot} + e^- \quad (3b)$$

$$M^{+\cdot} + mH \rightarrow m^{\cdot} + [M + H]^+ \quad (3c)$$

Quantities which may influence successful analyte ionization may include the ionization energy of the analyte and the homolytic bond strength of the matrix, $BDE(m—H)$, and the hydrogen atom affinity of the radical cation of the analyte, $HAA(^{+\cdot})$.

Possible mechanisms for the formation of $[M + Na]^+$ ions in MALDI

Mechanism 4: excited-state 'salt' chemistry.

$$mNa + h\nu \rightarrow mNa^* \quad (4a)$$

$$mNa^* + M \rightarrow m^- + [M + Na]^+ \quad (4b)$$

In this model, the matrix (mH) precipitates, at least in part, as the sodium salt, mNa. When mNa is excited, it can transfer Na^+ to the peptide, analogous to mechanism 1 for protonation.

Mechanism 5: protonation of the sodium salt of the analyte.

$$mH^*, mH^+ \text{ or } [mH + H]^+ + \{M - H + Na\} \rightarrow [\{MH + Na\} + H]^+ + \text{side product(s)} \quad (5)$$

Here, the mechanism for forming the sodium adduct is the same as for forming protonated molecules, because the $[M + Na]^+$ ion is not an Na^+ adduct of the analyte, but the protonated sodium salt of the analyte, $\{M - H + Na\}$. This would require that some portion of the analyte precipitate as the sodium salt.

Mechanism 6: gas-phase capture of Na^+ ions by the analyte.

$$Na^+(g) + M(g) \rightarrow [M + Na]^+(g)^* \quad (6a)$$

$$[M + Na]^{+*} \rightarrow [M + Na]^+ + h\nu \quad (6b)$$

$$[M + Na]^{+*} \rightarrow [M + Na]^+ + X^* \quad (6c)$$

This requires that $Na^+(g)$ is formed in the MALDI process, when NaCl is present in the matrix. This may be similar to processes observed in FAB, in which 'preionized' species are formed in high yield, because both desorption and ionization steps are not required—when the species desorbs it is already ionic (in this case, Na^+). Adduct formation as the bimolecular process shown in reaction (6a) yields an excited adduct, which can be stabilized by infrared emission of radiation in reaction (6b) or by collisional stabilization with a gas-phase molecule, X, in reaction (6c).[14]

Experimental observations and results

Laser power dependence of matrix and analyte ion formation. We have observed that the relative abundances of both matrix and analyte ions are power dependent. We decided to characterize this dependence, since it may provide insights into mechanistic aspects of MALDI. As energy available in the MALDI target increases, what may various mechanisms predict? If mechanism 2 were correct, we might detect only $mH^{+\cdot}$ at lowest laser powers, with mH_2^+ and MH^+ being formed when more energy is available. In contrast, if mechanism 1 is correct, there may be no correlation between the relative intensities of $mH^{+\cdot}$ and mH_2^+ as laser power varies.

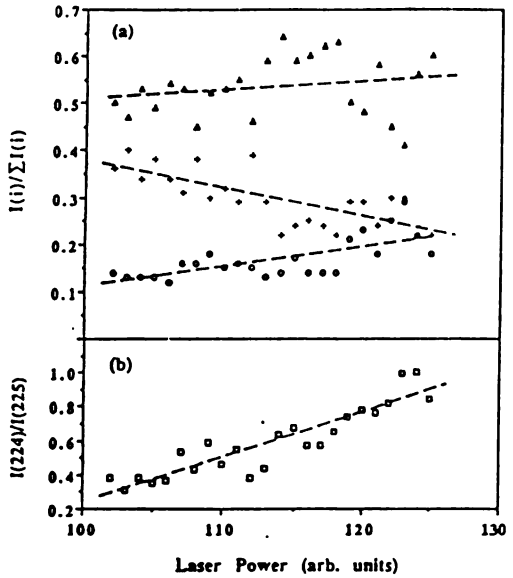


Figure 2. Variation of relative intensities of the peaks of matrix ions (sinapinic acid) as a function of laser power. In (a) relative intensities of the peaks of the three most intense matrix-related ions are shown: (+) $[SA + H]^+$ (*m/z* 225); (○) $SA^{+\cdot}$ (*m/z* 224); (△) $[SA + H - H_2O]^+$ (*m/z* 207). In (b), the ratio of the peak intensities representing the ionized and protonated matrix molecules is shown. In each case, lines representing simple linear fits to the data are shown.

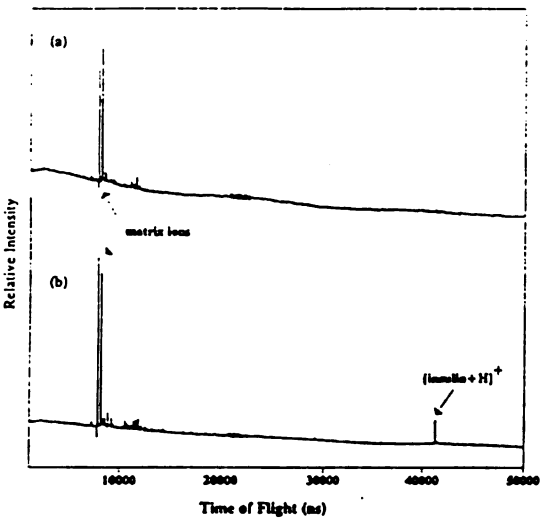


Figure 3. MALDI data obtained for the analyte insulin, using sinapinic acid as the matrix. In (a), at the laser power threshold for generation of ions, only matrix-related ions are observed. In (b), obtained from the same sample at higher laser power, matrix and analyte-related ions are observed. This shows that threshold for desorption/ionization is molecular mass dependent, at least in this case where the two molecules differ in molecular mass by a factor of more than 25.

When sinapinic acid, a matrix from which protonated analyte and matrix species are formed, is used, three ionic species relating to sinapinic acid (SA) are dominant: $[SA + H]^+$ (*m/z* 225), $SA^{+\cdot}$ (*m/z* 224) and $[SA + H - H_2O]^+$ (*m/z* 207). As in most laboratories in which MALDI is used, we do not measure actual power, but can systematically vary the laser power with a computer-controlled attenuator. The normalized data for the power dependence of the three matrix-related ions are given in Fig. 2(a). Obviously there is considerable scatter in the data, due to shot-to-shot fluctuations in the full N_2 laser power, variations in the target morphology, etc. In general, the fraction of the ion current representing $[SA + H]^+$ decreases as the laser power increases, that due to the ionized matrix molecule increases and that representing the dehydrated protonated matrix remains approximately constant. When the ratio of peak intensities for *m/z* 224; *m/z* 225 are plotted (Fig. 2(b)), the trend becomes clearer

Figure 2 shows how matrix ions vary with laser power. Is there a correlation between the relative abundance for $[M + H]^+$ for analytes with the relative abundance of any of the matrix ions, which may suggest a chemical 'link'? The answer is no, because it is well known that higher laser powers are required for generating ions of larger molecules such as insulin than for small matrix molecules.[15] This is shown in Fig. 3. Figure 3(a), obtained at low power, shows that only matrix ions are formed. At higher laser power, the spectrum shown in Fig. 3(b) is obtained; insulin present in the matrix is ionized and desorbed. This clearly demonstrates the fact that multiple processes are involved in the generation of gas-phase ions. Possibly, at lower power, protonation of larger analyte molecules can occur, but desorption cannot. Hence there is a correlation between analyte molecular mass and required laser power, with more power required to desorb large analytes. When that power is provided, protonated analytes appear in the gas phase.

Response factors for various analyte molecules. MALDI-MS is not generally used for the quantification of mixtures. When mixtures of peptides are encountered, such as in the analysis of tryptic digests, the signal intensity for each peptide of the mixture reflects, at least, the efficiencies of both the desorption and the ionization steps. Differential responses are often observed. Four peptides were selected to begin to characterize this, and a mixture of the four were analyzed using sinapinic acid and α-cyano-4-hydroxycinnamic acid as the matrices. The homopolyamino acids selected for studying the response factors for various peptides were hexaglycine (M_r = 360), hexaalanine (M_r = 444), pentaphenylalanine (M_r = 754) and hexatyrosine (M_r = 997). Figure 4(a) shows the MALDI spectrum of an equimolar mixture of these four peptides (10 pmol each) using sinapinic acid as the matrix. Only two of the four were clearly detected, pentaphenylalanine and hexatyrosine. A comparison of the *m/z* 300–500 region of the spectra shown with those for matrix alone suggests that there may be a small shoulder on the *m/z* 359 peak corresponding to *m/z* 361, and a small feature at *m/z* 445 as a shoulder on the *m/z* 449 peak. Clearly MALDI chemistry favors formation of ions of the larger two analytes of the mixture. The same is seen in Fig. 4(b), when the matrix is changed to α-cyano-4-hydroxycinnamic acid. Spectra

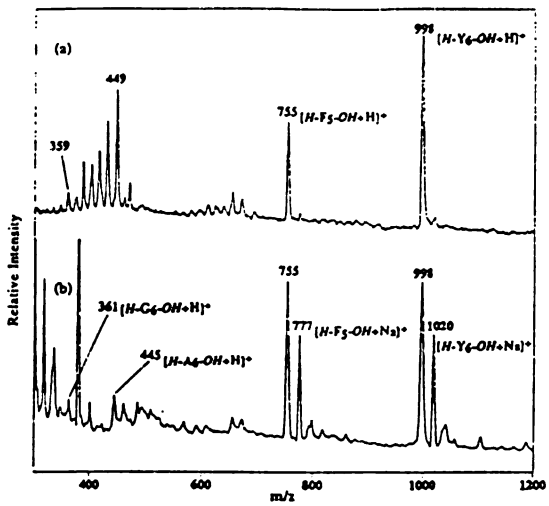


Figure 4. MALDI spectra of four-component mixture of peptides, using as the matrix (a) sinapinic acid and (b) α-cyano-4-hydroxycinnamic acid. If analyte ions are successfully desorbed and ionized, each will give peaks representing the protonated molecules. These will be detected at *m/z* 361 for hexaglycine, *m/z* 445 for hexaalanine, *m/z* 755 for pentaphenylalanine and *m/z* 998 for hexatyrosine.

features representing protonated hexaglycine and hexaalanine are much less intense than those for protonated hexaglycine and hexaalanine. Since the peaks at *m/z* 361 and 445 are more intense in Fig. 4(b), this may suggest that the protonating power of α-cyano-4-hydroxycinnamic acid is slightly higher than that for sinapinic acid. We note that the spectra in Fig. 4 were obtained using a laser power at which protonated hexatyrosine is observed. Thus, sufficient energy should be available for desorbing all of the smaller analytes. The difference in response for the smaller analytes of this set may not be due to desorption processes, but in the ionization step.

It has been suggested that the MALDI response for peptides correlates with their basicities;[16] however, the proton affinities of these peptides are not known. The proton affinity of hexaglycine is reported to be 236 kcal mole^{-1} [17] The proton affinities of the component amino acids are reported[18] to increase in the following order:

$$PA(\text{glycine}) < PA(\text{alanine}) < PA(\text{phenylalanine}) \sim PA(\text{tyrosine})$$
$$\sim 211 \quad < \sim 215 \quad < \sim 220 \quad < \sim 221 \text{ kcal mole}^{-1}$$

The proton affinities of the small peptides used might be expected roughly to parallel their component amino acids. Hence it appears as though there may be a correlation between the proton affinity of the analyte and the relative response.

To investigate the dependence of response on analyte characteristics, MALDI spectra of small molecules with known proton affinities were obtained, using three matrices. Selected molecules were chosen with a range of known proton affinities, ionization energies and bond strengths. With this set, there should be a correlation between response and analyte proton affinity if mechanism 1 or 2 is operative; if mechanism 3 is operative, the best response may be expected for analytes that have a low ionization energy and a high hydrogen atom affinity, $BDE(M^+—H)$. Also, these molecules have molecular masses such that, if protonated, would yield peaks at *m/z* values that do not correspond to matrix-related peaks. Finally, analytes were selected that would be soluble in the solvents used for delivery of the matrix to the probe tip. In this experiment, water–acetonitrile (2:1) was used as the solvent. The compounds chosen are listed in Table 1. The six compounds were characterized in the three matrices sinapinic acid, α-cyano-4-hydroxycinnamic acid and dihydroxybenzoic acid. In all three matrices, the results were the same. Signals were observed representing the protonated analytes for all of the compounds in Table 1 except 4-cyanobenzaldehyde and thymidine. It is noteworthy that direct laser desorption without matrix can produce ions from these molecules only with much higher laser power than the power threshold in MALDI, so these results were obtained at laser powers at which true matrix-assisted processes occur. Based on the data in Table 1, it appears that the analyte must have a proton affinity that lies above some value, $208 < X < 224$ kcal mol^{-1}, to be detected (in protonated form). The ionization energy data in Table 1 suggests no correlation of analyte ionization energy with MALDI detectability.

Table 1. Compounds used for studying the relationship between response factors and thermochemical parameters[a]

Compound	M_r	PA[b]	IE[c]	HAA[d]	Signal observed?
4-Cyanobenzaldehyde	131	187	229	102	No
Thymidine	242	208	196	90	No
Phenazine	180	224	192	102	Yes
Acridine	179	232	180	98	Yes
1,7-Diaminoheptane	130	238	198	122	Yes
1,8-Bis(dimethylamino)naphthalene	214	242	149	77	Yes

[a] See appendix.
[b] *PA* = proton affinity (kcal mol^{-1}).
[c] *IE* = ionization energy (kcal mol^{-1}).
[d] *HAA* = hydrogen atom affinity of the ionized molecule, $BDE(M^{+\cdot}—H)$ (kcal mol^{-1}).

The same is true for the hydrogen atom affinities of the ionized molecules. Hence the 'protonating powers' of each of the three matrices selected are similar.

Formation of protonated vs sodiated peptides. As shown in Fig. 1(a), (b) and (c), the formation of $[M + H]^+$ vs $[M + Na]^+$ ions is highly matrix dependent for hexatyrosine. One of the possible explanations as to why only $[M + H]^+$ ions, but no $[M + Na]^+$ ions, were observed in Fig. 1(a) for the sinapinic acid matrix is that Na^+ may not be available for the sodiation of hexatyrosine because of the form it takes within the sinapinic acid matrix. Indeed, neither Na^+ nor $[SA + Na]^+$ ions were observed at the threshold laser power at which protonated hexatyrosine was generated when sinapinic acid was used as matrix with adding excess amount of NaCl (Fig. 1(a)). The matrix dependence of the abundances of five ions (Na^+, $[matrix + H]^+$, $[matrix + Na]^+$, $[hexatyrosine + H]^+$ and $[hexatyrosine + Na]^+$) obtained from the spectra which yield Fig. 1(a), (b) and (c) are shown in Fig. 5. The coincidence of Na^+, $[matrix + Na]^+$ and $[hexatyrosine + Na]^+$ ions is obvious. Matrices that yield intense Na^+ signals also yield intense $[M + Na]^+$ and $[mH + Na]^+$ signals.

To characterize further any correlation between sodiated and protonated species, the power dependence of various ions was studied for hexatyrosine in α-cyano-4-hydroxycinnamic acid, the matrix which yields both protonated and sodiated species. Results are summarized in Fig. 6. Figure 6(a) shows that the ratio of sodiated to protonated species of both analyte and matrix correlate. Conditions that favor formation of sodiated matrix also favor formation of sodiated analyte. In Fig. 6(b), the power dependence of the Na^+ signal is shown. This correlates well with the data shown in Fig. 6(a). In particular, when experimental conditions yielded lower Na^+ abundances (indicated as points A and B), there are also low points in the Fig. 6(a) data, showing a strong correlation between the abundance of gas-phase sodium ions and sodiated molecules. We also note that, when multiple samples were characterized at a variety of laser powers, there is a consistent correlation between the $I(XNa^+)/I(XH^+)$ and $I(Na^+)$.

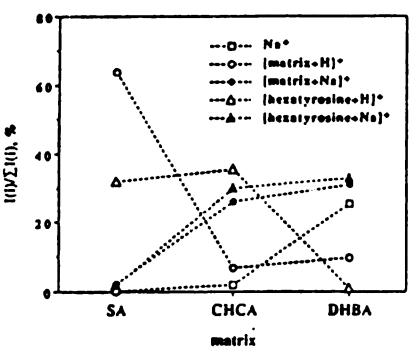


Figure 5. Relative abundances of sodiated and protonated matrix and analyte ions in the three matrices studied. SA = sinapinic acid; CHCA = α-cyano-4-hydroxycinnamic acid; DHBA = 2,5-dihydroxybenzoic acid. The data show that sodiated species are formed from matrices from which Na^+ ions are also desorbed.

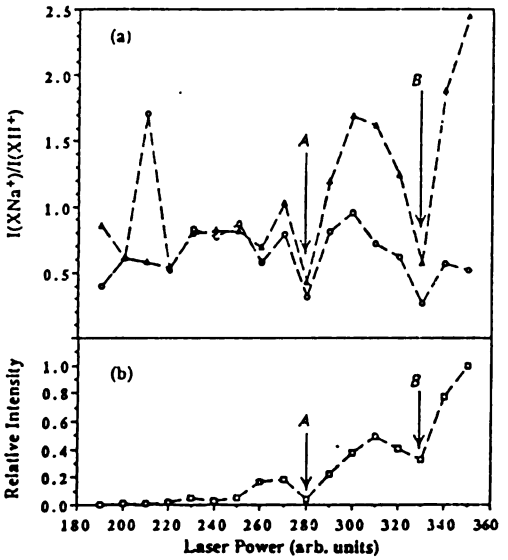


Figure 6. Power dependence of various ions for hexatyrosine in α-cyano-4-hydroxycinnamic acid. In (a), the power dependence of peak intensity ratios of sodiated-to-protonated species is shown; (○) X = matrix; (△) X = hexatyrosine. In (b), the power dependence of the peak intensity representing Na^+ ions is shown. When experimental conditions yielded lower Na^+ abundances (indicated as points A and B), there are also low points in (a), showing a strong correlation between the abundance of gas-phase sodium ions and sodiated molecules.

The results in Fig. 1, obtained with extreme amounts of NaCl added in each, paralleled our initial observations without NaCl addition. The matrix 2,5-dihydroxybenzoic acid tends to generate sodiated molecules, even when no NaCl is added. Do small NaCl impurities result in dominant $[M + Na]^+$ peaks when some matrices are used (as supplied)? To study this, a sample of α-cyano-4-hydroxycinnamic acid was desalted using a cation-exchange resin.[9] When this matrix is desalted, and used in the MALDI analysis of (desalted) hexatyrosine, the spectrum shown in Fig. 7 is obtained. Note

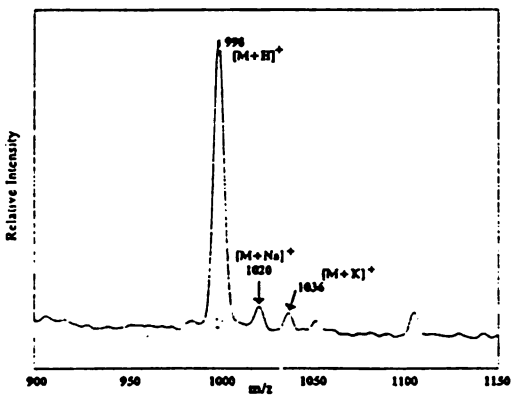


Figure 7. Portion of the MALDI mass spectrum of purified hexatyrosine, using a desalted α-cyano-4-hydroxycinnamic acid matrix. A comparison with the data in Fig. 1(b) shows that this matrix can generate protonated analyte molecules when the NaCl content is reduced.

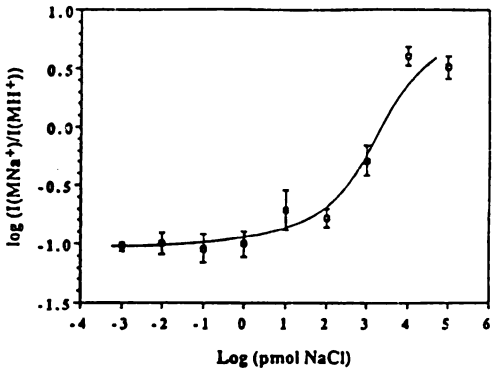


Figure 8. Titration curve for the addition of NaCl to 40 pmol of hexatyrosine and 6000 pmol of α-cyano-4-hydroxycinnamic acid. The data show that very small amounts of NaCl do not result in the formation of $[M + Na]^+$ ions. Larger amounts, similar to the molar amounts of matrix used, are required.

that the $[M + H]^+$ peak now dominates the spectrum. When only the analyte is desalted, sodiated ions are formed. This shows that NaCl impurities in the matrix were responsible for the observed ions.

With pure matrix material, how much NaCl is required to generate the sodiated ions? More specifically, are very small amounts required, comparable to the (molar) amount of analyte used, or are larger amounts required, comparable to that for the matrix? Is there a stoichiometric optimum that might suggest that the Na^+ adducts originate from either the salt of the matrix or the salt of the analyte? The desalted matrix–analyte mixture was titrated with NaCl, and MALDI spectra were obtained. Figure 8 shows data for the ratio of the $[M + Na]^+$ peak intensity to the $[M + H]^+$ peak intensity as various amounts of NaCl are added to the matrix and sample (for a constant analyte-to-matrix ratio of 1:150). It is found that the relative $[M + Na]^+$ ion yield increases dramatically when the NaCl-to-α-cyano-4-hydroxycinnamic acid molar ratio is]1:3, at very large NaCl-to-analyte molar ratios.

Since large amounts of NaCl are required to yield sodiated analytes, any of the three mechanisms proposed could be operative. If mechanism 5 is operative, there is a simple test. It was observed that pentaphenylalanine forms $[M + Na]^+$ ions in MALDI. If it is really an $[M - H + Na]H^+$ ion (protonated sodium salt of the *C*-terminus), then the methyl ester of pentaphenylalanine should not form a sodium adduct, since precipitation of the peptide as the sodium salt ($RCOO^-Na^+$) cannot occur. The methyl ester of pentaphenylalanine, H-FFFFF-OMe was made and analyzed by MALDI. Figure 9(a) and (b) show the resulting spectra using α-cyano-4-hydroxycinnamic acid as the matrix without and with addition of excess NaCl to the sample, respectively. (Note that, from the data presented in Fig. 1, the matrix which was selected is one from which both protonated and sodiated ions can evolve.) It is evident that sodiation does take place even when the analyte cannot precipitate as the sodium salt. This conclusion parallels the observation of Wan *et al.*[19]

We note an additional observation, not for $[M + Na]^+$ ions, but for + 1 forms of the analyte con-

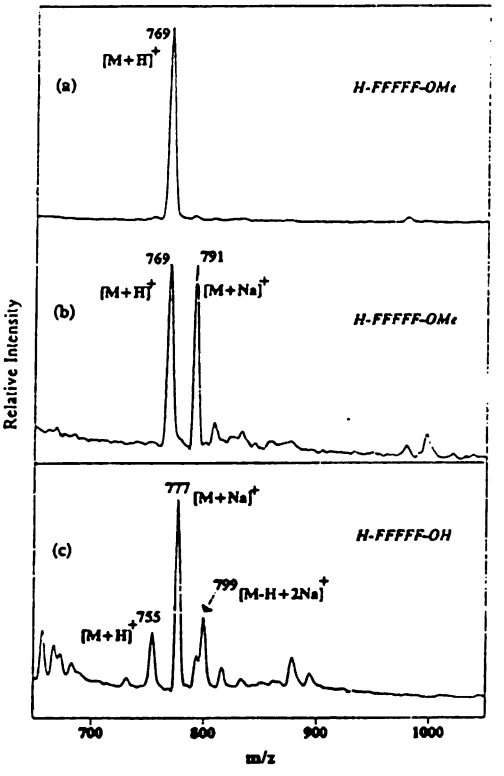


Figure 9. Portions of the MALDI mass spectra for (a) the methyl ester of pentaphenylalanine, using pure α-cyano-4-hydroxycinnamic acid matrix; (b) the same sample as in (a), with excess NaCl added; (c) pentaphenylalanine, with α-cyano-4-hydroxycinnamic acid and excess NaCl. The amount of NaCl added to yield the spectra (b) and (c) is the same as that used to obtain the data in Fig. 1.

taining multiple sodium atoms. Figure 9(c) shows the pseudomolecular ion region for pentaphenylalanine using α-cyano-4-hydroxycinnamic acid with NaCl. In addition to observing an $[M + Na]^+$ ion, we also observe a species containing two sodiums, $[M - H + 2Na]^+$. Note that $[M - H + 2Na]^+$ ions are not observed in the spectrum of the methyl ester of pentaphenylalanine when NaCl is present (see Fig. 9(b)). These observations suggest that, whereas the $[M + Na]^+$ species may be simple Na^+ adducts of the free acid form of the peptide molecules, the $[M - H + 2Na]^+$ ions may well be Na^+ adducts of the sodium salt of analyte, more accurately designated as $[M - H + Na]Na^+$.

Comments on possible ionization mechanisms

The focus of the discussion that follows is on the chemistry of the ionization process of MALDI, not the desorption step. Models for desorption, more accurately described as sublimation, ablation or volatilization,[20] have been proposed which seem to explain well the dynamics of the system in this regard. We evaluate here the proposed mechanisms in the light of the experiments discussed here, and the prior literature.

Mechanism 1: excited-state acid–base chemistry. This mechanism based on the excited-state acid–base chemistry has been proposed and discussed by Russell and co-workers,[21] who pointed out that sinapinic acid methyl ester is also a good matrix, so it is not the acid group that is responsible for the ionization step. The mechanism was proposed without commitment as to whether it occurs in the gas phase or in the condensed phase. This mechanism is consistent with the data in Table 1. Those analyte molecules with high proton affinities tend to form abundant analyte ions, $[M + H]^+$. The weakness of this mechanism is that it fails to explain the formation of the radical cation of sinapinic acid in the MALDI experiment. However, the data in Fig. 2 may suggest that formation of $mH^{+\cdot}$ may be independent of the processes that lead to protonated analyte molecules. Also, wherever the process occurs, it requires a separation of charge—work must be performed against attractive coulombic forces. However, this is true in all desorption/ionization experiments where neutral targets yields ions—both positively and negatively charged species must be formed and separated at some point.

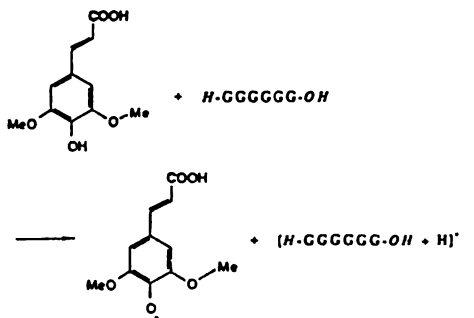
It has been reported that, in solution, molecules similar to matrix molecules, aromatic acids, tend to show increased pK_a values when they are electronically excited. In contrast, the molecules with hydroxyl functional groups tend to show a decrease in pK_a when they are excited.[22] Although these data are for solutions, the concepts should still hold some validity. The three matrices studied here all have carboxylic acid and hydroxyl functional groups. The increased acidity of matrix molecules while they are excited should involve the hydroxyl group. However, the pK_a of the phenol molecule is decreased from 9.85 to only 3.01 in its excited state (in comparison, the pK_a of ground-state benzoic acid is 4.20). In this increase of acidity high enough to initiate a proton transfer reaction? If the acidity of the matrix molecule is crucial for the proton transfer from matrix molecules to analyte molecules, then some highly acidic molecules should work equally well as matrices even when they are in their ground states.

One of the strongest arguments for mechanism 1 is that some spectra have been reported where, at threshold laser powers, MALDI generates protonated analyte molecules without forming matrix-related ions.[23] If MALDI can be performed where only analyte ions are formed, then this mechanism would be one of the few ways to describe the system. However, if this is the mechanism, then why do not more molecules work well as MALDI matrices? Also, if this is the mechanism, does it occur in the gas phase or in the condensed phase? Certainly, if electronically excited matrix molecules are involved, their concentrations are highest when they are first formed—in the condensed phase. Parker and Hercules[24] discussed proton transfer reactions via excited molecules in laser desorption (no matrix), using a mechanism that is certainly relevant to this discussion. They proposed the desorption of dimers, with excited-state lifetimes on the order of microseconds—the excimers separate after desorption, in the gas phase, to yield positive and negative ions. In MALDI, there could be gas-phase (matrix*) (analyte) complexes which are either desorbed intact, or formed in a collision, which could participate in proton transfer, with subsequent separation of the ion pair. Parker and Hercules recognized that ion-pair separation requires high energies, but they suggested that enough is available in laser desorption to allow the process to occur. The many experiments that have been reported on energy spreads of MALDI-generated ions, which suggest that all ions have a constant velocity spread, not a constant energy spread, certainly suggest that many collisions occur in the volume above the target. Gas-phase collisions could deposit additional energy, facilitating ion-pair separation. Clearly, more work remains to correlate positive- and negative-ion spectra before such mechanisms can be ruled out. The major advantage of a condensed-phase version of mechanism 1 is that the condensed-phase matrix molecules can 'solvate' the $[matrix - H]^-$ anion while the protonated analyte desorbs. If this is the mechanism, then the dominance of singly protonated peptides remains to be explained.

The thermochemical aspects of mechanism 1 were also considered. Owing to the difficulty of obtaining thermochemical data for excited-state matrix molecules, we consider first the ground-state acid–base chemistry of gas-phase species which is similar to mechanism 1.

$$mH + M \rightarrow m^- + [M + H]^+ \qquad (7)$$

ΔH for equation (7) can be estimated using two relevant quantities: gas-phase acidity, ΔH_{acid} of the matrix, which is $\Delta H(mH \rightarrow m^- + H^+)$, and proton affinity of the analyte, which is $-\Delta H(M + H^+ \rightarrow [M + H]^+)$. The thermochemical data for the sinapinic acid matrix molecules are not available, and model compounds must be considered as analogues. Five molecular analogues, selected to represent the various substructural features of sinapinic acid, with their ΔH_{acid} values,[25] are presented in Fig. 10. The ΔH_{acid} data show that generation of an anionic site on oxygen atoms requires less energy than that on the carbon atoms. It also shows that ΔH_{acid} values for carboxylic acid and hydroxy groups are similar in the gas phase. Since MALDI still works for the methyl ester of sinapinic acid as the matrix,[21] we shall consider the role of the hydroxy group here. The overall ΔH of Eqn (8) is approximated to be 114 kcal mol^{-1} for hexaglycine (using ΔH_{acid}(sinapinic acid) = 350 kcal mol^{-1}; proton affinity (hexaglycine) = 236 kcal mol^{-1} (Ref. 17)). Hexaglycine is chosen for this discussion since its proton affinity is known.



(8)

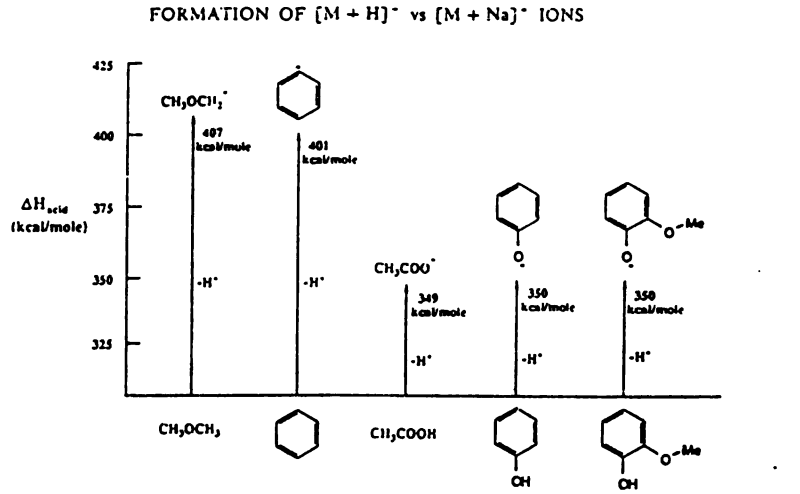


Figure 10. Gas-phase acidity data for molecules representing the various substructural features of sinapinic acid.

The proton transfer from the ground state of the sinapinic acid matrix molecule to this polypeptide analyte is endothermic by 114 kcal mol^{-1}. From Table 1, we know phenazine can give a signal in the MALDI experiment, although the proton affinity of phenazine is smaller than that for hexaglycine. The proton affinity of phenazine is 224 kcal mol^{-1}. ΔH for proton transfer from a sinapinic acid matrix molecule to phenazine is calculated to be 126 kcal mol^{-1}. That is, the proton transfer from the ground-state sinapinic acid matrix molecule to a phenazine analyte is endothermic by 126 kcal mol^{-1} (5.46 eV).

Bond strengths or ΔH_{acid} values are not available for the excited state of sinapinic acid matrix molecules, but the changes between excited state and ground state cannot differ by more than the energy deposited upon excitation. The energy of a 337 nm photon is 3.96 eV, and the proton transfer from sinapinic acid to phenazine is endothermic by 5.46 eV. Either a two-photon excitation of a single sinapinic acid matrix molecule is needed to make proton transfer from sinapinic acid to phenazine exothermic, or it can be done with one-photon excitation where ΔH_{acid}(sinapinic acid) is lower in the condensed phase because of 'solvation' of the resulting [sinapinic acid − H]⁻ anion. This situation is illustrated in Fig. 11. Mechanism 1 could occur by one-photon excitation if the solvation energy of the [sinapinic acid − H]⁻ ion in surrounding matrix molecules is greater than 1.5 eV. At least in the gas phase, solvation energies for anions can be considerable. An example is $OH^- + H_2O \rightarrow HO^- \cdots H{-}OH$, $\Delta H = 23.9$ kcal $mol^{-1} \approx 1.1$ eV. If the resulting [sinapinic acid − H]⁻ anion can be solvated by several surrounding matrix molecules, the solvation energy could be greater than 1.5 eV and allow mechanism 1 to occur by one-photon excitation. From Table 1, we also know that thymidine does not give a signal in the MALDI experiment. The proton affinity of thymidine is 208 kcal mol^{-1}, so the proton transfer from the ground-state sinapinic acid matrix molecule to a thymidine analyte is endothermic by 142 kcal mol^{-1}, or 6.16 eV. If mechanism 1 is operative and is through a one-photon excitation, then a reasonable solvation energy of the

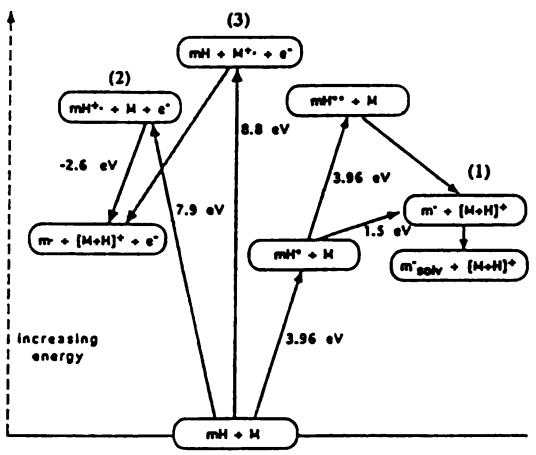


Figure 11. Summary of thermochemical considerations for the formation of protonated peptides in MALDI. M = hexaglycine, mH = sinapinic acid matrix and the numbers used to label various pathways correlate with the numbering scheme used for the mechanisms discussed. A 337 nm photon has an energy of 3.96 eV.

[sinapinic acid − H]⁻ anion in surrounding sinapinic acid matrix molecules would be between 1.5 and 2.2 eV. Thus, for a one-photon excited-state acid–base chemistry mechanism, proton transfer must occur in the condensed phase. However, consider the data shown for the laser power dependence of analyte peaks; there is a sharp irradiance threshold for the MALDI process.[20] We can form [SA + H]⁺ at lower laser powers than [insulin + H]⁺ and PA(sinapinic acid) is less than PA(insulin), so proton transfer chemistry can occur even at lower laser powers where vaporization of large analytes cannot. Since vaporization is a multi-photon event, the chemistry can certainly also be so. Finally, while lasers allow for multi-photon processes, in this case two-photon excitation of a single matrix molecule deposits an energy which exceeds the ionization energy, making the mechanisms shown in Fig. 11 via mH** unlikely.

Mechanism 2: chemistry following matrix photoionization. Matrix photoionization, followed by ion–molecule reactions, has been suggested by Hillenkamp and co-workers,[19,26] Mowry and Johnston.[27] Karas[28] and Beavis and co-workers,[23,29] although again, the process could involve gas-phase and/or condensed-phase species. We propose two versions. Mechanism 2 relies on photoionization as the first step, followed by proton transfer. Mechanism 2' uses an additional step, with the protonated matrix molecule protonating the analyte—the true analogue to chemical ionization (CI). Mechanism 2' should be considered because MALDI, FAB and CI have the common feature of large 'matrix-to-analyte' ratios, which in CI is the reagent gas-to-analyte ratio. It has been suggested that, in many FAB analyses, the experiment is essentially a glycerol CI experiment.[30] Like CI experiments, 'reagent ions' and their products are usually detected in the same spectrum.

This mechanism is the simplest. Molecular ions *are* formed by photoionization in laser desorption experiments, so we should not be surprised if this is how the chemistry starts in the MALDI experiment.[26] There are several strengths of such mechanisms. First, when MALDI-generated spectra are obtained with sufficient resolution, we see for typical compounds both $M^{+\cdot}$ and MH^+ for the matrices, but only protonated analyte molecules,[31–37] which would be consistent with this mechanism. Second, it is consistent with the data in Table 1, in that those analyte molecules with high proton affinities tend to form analyte ions, $[M + H]^+$. Third, photoionization is surely a multi-photon process, as is the MALDI process.[20,38]

However, the data presented here suggest that, although molecular ions are protonated molecules are formed, they may involve two distinct processes that need not be chemically linked. At higher powers, molecular ions can be formed, but they are not necessarily required for the formation of the other ions observed The results presented could be interpreted to indicate that photoionization is simply a process with a higher threshold than the dominant processes in MALDI. There have been experiments described in which protonated analyte ions are formed but no matrix ions are formed (at threshold), which could negate such mechanisms. It has been suggested that photoionization does initiate the process, but in the high-pressure neutral plume that is formed, analytes act as proton traps which can completely scavenge protons, such that the matrix ions 'disappear'.[26] In the light of the literature on the use of high-pressure mass spectrometry to measure equilibrium constants, such an explanation seems unlikely.

To understand entirely the protonation mechanism of MALDI, one important question that has to be answered is whether the ionization process involves gas-phase or condensed-phase chemistry. For a gas-phase mechanism, neutral molecules must be desorbed in the MALDI experiment. This has been shown.[27] Reported experiments involving time-delayed ion extraction from MALDI ion sources[19,28] show that such delays enhance the abundance of protonated analyte molecules. These results strongly suggest that gas-phase chemistry can contribute to the ions observed, at least in time-delayed extraction experiments. While suggesting a possibility, they do not prove that the same process could occur in a 30 kV MALDI source under continuous extraction conditions.

In the condensed phase, where exceedingly large peptides can be surrounded by hundreds of matrix molecules undergoing irradiation, one may well expect multiple protonation processes. In fact, since the process begins in the condensed phase, it may not have been surprising if MALDI spectra looked more like electrospray data than CI spectra. The fact that most of the analyte ions are singly protonated might be best correlated with gas-phase chemistry, where analytes, once protonated, cannot encounter another protonating reagent ion owing to coulombic interactions. However, two aspects of the experiment would suggest that gas-phase ion–molecule reactions do not, under typical MALDI conditions, yield protonated analyte molecules. The first is the high voltage, up to 30 kV, used in many MALDI ion sources. Under such conditions, ions are quickly accelerated to kinetic energies where collision-induced dissociation processes occur, not ion–molecule reactions such as proton transfer[39] However, suppose ion–molecule reactions can occur in the MALDI ion source. If these are responsible for the analyte ions formed, the ratio of matrix ions to analyte ions is smaller than in CI, suggesting extensive conversion of reactant ions to product ions, indicative of a large number of collisions. What other indicators are there of a high collision frequency? In FAB, protonated glycerol molecules cluster with subsequent glycerol molecules in the gas phase to form $[G_n + H]^+$ ions with m/z values exceeding 1000. Surely, with polar molecules such as those used as MALDI matrices, similar cluster ions would be expected if a large number of gas-phase collisions occur for the ions initially formed. Thus, although collisions do appear to occur in the desorbed plume, the absence of $(mH)_nH^+$ ions suggests that the collisions occur at energies at which ion–molecule reactions are not favored.

If one considers CI, FAB and MALDI spectra, one other difference may be relevant. In CI and FAB experiments, fragmentation follows protonation. Even when small analytes are used in MALDI, prompt fragmentation is rarely observed. This might be consistent with a condensed-phase mechanism where energy within the solid is rapidly dissipated following protonation, leading to desorbed protonated analyte molecules that do not contain sufficient internal energy to fragment. Yet another difference between CI, FAB and MALDI is the specific, but as yet poorly defined, requirement for a successful matrix. It is puzzling that so few compounds are effective MALDI matrices,[20] while many compounds can be used in FAB matrices. However, this observation may not be related to the chemical step of protonation, but to processes related to energy transfer and subsequent desorption of analytes.

Further, it should be noted that such conclusions are directed toward the typical MALDI time-of-flight (TOF) experiment. In MALDI-Fourier transform mass spectrometry (FTMS), where ions may be detected tens to hundreds of milliseconds after the initiation of the MALDI process, ion–molecule reactions could certainly contribute to the species observed. However, the situation is very different in MALDI-TOF, where ions are

detec
MAL
descr
for th
situat
spect
extra
descr
lowe
W
ent i
the g
on th
discu
contr
throu
or an

Mecl
the p
proto
ioniz
and v
MAL
tions
of a
How
spect
not b
bility
energ
mole
analy
analy
detec
is ce
extra
the c
T
form
syste

detected on the microsecond time-scale. Also, if MALDI is performed in a simple FTMS 'cell,' ions and desorbed neutrals may be present in the same volume for the entire ion formation and analyses processes. The situation is very different in more conventional mass spectrometers, in which ionic species are quickly extracted from the source where the pressure of desorbed neutrals is highest, into a flight tube of much lower pressure.

While the energetic aspects of mechanism 2 are different if the chemistry occurs in the condensed phase or in the gas phase, we can make some qualitative comments on the thermochemistry. These will be included in the discussion of mechanism 3, where we can compare and contrast the formation of protonated analyte molecules through the first step of either matrix photoionization or analye photoionization.

Mechanism 3: chemistry following analyte photoionization. In the previous mechanisms, it is assumed that the matrix protonates analyte molecules, and that activated or ionized matrix molecules chemically lead to protonated analyte molecules. Even though many analytes in MALDI do absorb at 337 nm, the relative concentrations of matrix and analyte make direct photoactivation of an analyte unlikely in the MALDI experiment.[12] However, care must be taken to realize that all of the species represented in a MALDI mass spectrum need not be formed via the same processes. One such possibility is suggested in mechanism 3. In this mechanism, energy is accumulated from multiple excited matrix molecules (as it is to induce desorption) to ionize the analyte, and though this intermediate, protonated analytes evolve. If this were the case, we might expect to detect ionized analytes (odd-electron ions), although it is certainly reasonable that the ionized analytes could extract a hydrogen atom as they make their way out of the condensed phase.

This mechanism does not require that matrix ions be formed, and thus is consistent with those reports of systems in which spectra were obtained containing only analyte ion peaks, free of matrix peaks. There are cases where, in MALDI experiments on oligometallocenes such as polymeric compounds containing ferrocene, molecular ions, not protonated molecules, are formed.[40] In these experiments, the matrix does play a role. It may assist in desorption alone, or may be involved in charge-transfer processes following matrix photoionization. A third possible explanation for such cases where MALDI leads to odd-electron molecular ions of the analyte is similar to that found in the CI literature for organometallic compounds. In some cases, molecular ions, not protonated molecules, are formed in methane CI. It has been suggested that the first step is proton transfer, but the protonated molecule rapidly fragments by H atom loss, leading to the observed molecular ion:

$$CH_5 + M \rightarrow [MH]^{+\bullet} \rightarrow M^{+\cdot} + H^{\cdot} \qquad (9)$$

If mechanism 3 is operative, the analyte which has a low ionization energy and, once ionized, has a high hydrogen atom affinity, should have a high tendency to form $[M + H]^+$ in MALDI. From Table 1, the trend shown in the ionization energies and hydrogen atom affinities does not correlate well with mechanism 3. The data presented here most clearly indicate that the property most associated with analyte detectability is proton affinity.

To put the first three mechanisms into perspective in terms of their energetic accessibility, some thermochemical aspects of the photoionization mechanisms (2 and 3) should be acknowledged, to be contrasted with those presented in the discussion for mechanism 1. Mechanisms 2 and 3 are initiated by photoionization processes. The photoionized molecules are matrix and analyte molecules in steps (2a) and (3b), respectively. The situation where the matrix is sinapinic acid and the analyte is a polypeptide will be taken as an example for the following discussions. Although the ionization energy of sinapinic acid has not been reported, it can be estimated from its molecular analogues,[25] some of which are summarized in Fig. 12. It appears that the lowest energy form of the radical cation of sinapinic

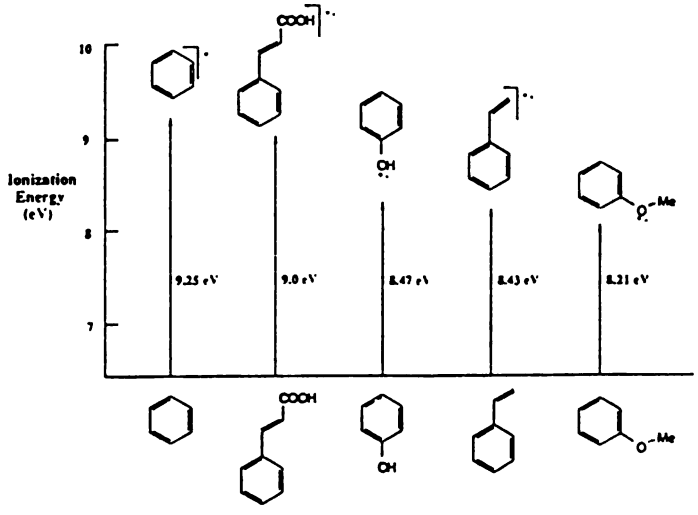
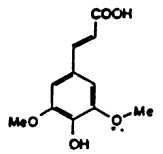


Figure 12. Gas-phase ionization energy data for molecules representing the various substructural features of sinapinic acid.

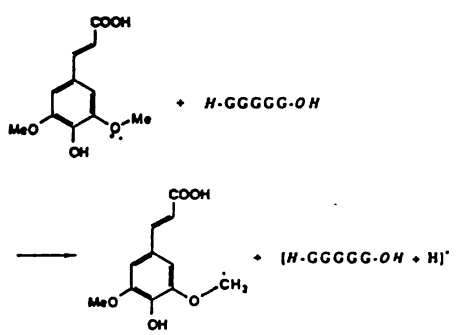
acid is, most likely, that shown as:



(I)

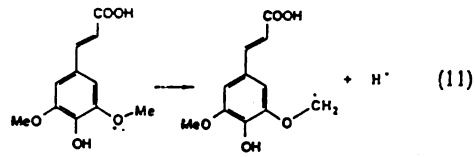
with the charge delocalized somewhat throughout the conjugated system. The ionization energy of sinapinic acid is estimated to be close to 8.2 eV, the ionization energy of methoxybenzene. Comparing *IE*(phenol), 8.47 eV, and *IE*(1,2-dihydroxybenzene), 8.15 eV,[25] it is found that the *o*-hydroxy group can influence the ionization energy by lowering the value by approximately 0.3 eV. This refines our estimate of the *IE*(sinapinic acid) to 7.9 eV. The ionization energies of these molecular analogues which were used for the estimation were measured in the gas phase. The ionization energy of a sinapinic acid molecule in the condensed phase should be lower than this value.

Using a value for *IE*(sinapinic acid) of 7.9 eV, three comments can be made. First, it is lower than *IE*(polypeptide). Again, although the ionization energy of a polypeptide such as glycine has not been reported, it can be estimated by its molecular analogues. Glycine and acetamide, which are analogous to the terminal and amide groups of a polypeptide chain, respectively, have ionization energies of 8.8 and 9.65 eV.[25] Compared with a polypeptide, sinapinic acid has a lower ionization energy. This suggests that the photoionization mechanism begins with photoionized sinapinic acid molecules, not polypeptide ions. Mechanism 2, rather than mechanism 3, is more likely to be operative in the MALDI source. Second, the photoionization of sinapinic acid is a two-photon process. Further, the photoionization process of a polypeptide molecule requires more than two 337 nm photons, which is less likely to be a dominant process in this case. Third, our analysis suggests consideration of a specific protonation reaction. The protonation reaction that will be considered here is shown in reaction (10).



(10)

The data presented on the ionization energies of matrix and analyte molecules suggest that mechanism 2 is favored over mechanism 3. Is the proton transfer described in Eqn (10) thermochemically favored? Equation (10) can be separated into two processes, Eqns (11) and (12). It is assumed that the proton transfers from the methoxy group of sinapinic acid to the analyte if the ionization site for sinapinic acid occurs in the methoxy group.



(11)

$$\text{polypeptide} + H^+ \rightarrow [\text{polypeptide} + H]^+ \quad (12)$$

The enthalpy change, ΔH, for the two equations can be evaluated separately. ΔH for Eqn (11) is not known, and will be approximated using ΔH of the equation[25]

$$CH_3OCH_3^{+\cdot} \rightarrow CH_3OCH_2^{\cdot} + H^+ \quad \Delta H$$

$$= 175.5 \text{ kcal mol}^{-1} \quad (13)$$

ΔH of Eqn (12) is the proton affinity of the polypeptide. The proton affinity of hexaglycine has been reported to be 236 kcal mol^{-1},[7] which can be used to estimate the minimum value for the proton affinity in Eqn (12). Thus, the overall ΔH for the proton transfer described in Eqn (10) is exothermic with a maximum value of -60 kcal mol^{-1}. The overall information for thermochemical aspects of mechanisms 2 and 3 is summarized in Fig. 11.

Two conclusions can be reached: (i) a photoionization mechanism would begin with a photoionized sinapinic acid matrix molecule, not an ionized polypeptide; (2) the radical cation of sinapinic acid can protonate a polypeptide in an exothermic process. These conclusions are consistent with (1) the observation of the radical cation of sinapinic acid in MALDI mass spectra, although more radical cations are observed at higher laser power, and (2) the fact that the methyl ester of sinapinic acid still works well as a matrix.[21]

Obviously the thermochemical data presented here are gas-phase data, and variations due to the condensed phase cannot be rigorously considered. However, the summary in Fig. 11 carries some important concepts. Mechanism 1 should be thermodynamically favored over mechanisms 2 and 3 since excitation should require less energy than ionization, even in the solid state. In terms of the energetics of the overall process, mechanism 1 is also preferred owing to stabilization of deprotonated matrix molecules in the condensed phase. Obviously, additional considerations, such as the fate of the electron in mechanisms 2 and 3, could also influence the overall thermochemistry. However, thermochemical considerations alone would favor mechanism 1 for the generation of protonated analyte molecules.

Mechanisms for forming sodium ion adducts of analyte molecules. A variety of aspects of the experiment should be dealt with in attempting to evaluate mechanism 4–6. Obviously, they depend on the the precipitation process and the nature of the solid formed. To what extent does a peptide precipitate from a solution containing NaCl as the neutral peptide and as the sodium salt? This is a critical part of the analysis, although the realities of the experiment do not allow for generalizations based on equilibrium values such as solubility products, K_{sp}. This is becaue solvent systems are selected so they can be

rapidly evaporated, and precipitates are not formed slowly, approaching equilibrium conditions. A second aspect is what appears to be the 'condition' between the formation of protonated and sodiated analytes. While we cannot address these aspects here, we can begin to consider the mechanisms through which sodiated ions may be formed.

Mechanisms 4–6 will be treated together here. There is no extensive prior literature in related fields on which to rely such as the CI literature. Sodiated ions are formed in FAB and also in MALDI, although it is not known if the FAB mechanism is gas phase or if $[M + Na]^+$ ions desorb intact from solution. We do know, however, that alkali metal ions will form complexes with polar molecules including peptides, from the potassium ion ionization of desorbed species (K^+IDS) literature.[39,41,42] In such experiments, K^+ and Na^+ have been reported to complex with a large variety of desorbed molecules, without inducing any fragmentation. In K^+IDS, adduct formation occurs in the gas phase. Hillenkamp and co-workers[19] also showed that Na^+ ions could be injected into a MALDI source, and analyte adduct ions could be formed via gas-phase processes (again, in a field-free situation). We have shown that, as the kinetic energy of alkali metal ions increases, the cross-section for adduct formation rapidly decreases.[39] Thus, a gas-phase mechanism may be less likely in MALDI sources that use high voltages, such as the instrument used to obtain the data presented here.

In mechanism 4, excited-state 'salt' chemistry is suggested. While this mechanism parallels mechanism 1 for protonation, and the amount of salt required for the formation of sodium ion adducts is of the same order of magnitude as the molar amount of matrix used, there are problems with this mechanism, largely thermochemical in nature. Consider the processes (14) and (15) to contrast mechanisms 1 and 4:

$$mH + M \rightarrow [M + H]^+ + m^- \qquad (14)$$

$$mNa + M \rightarrow [M + Na]^+ + m^- \qquad (15)$$

These two processes contribute to the overall reaction enthalpy. First, the mH or mNa bond must be heterolytically cleaved. The bond dissociation energies that are relevant are $BDE(m^-—H^+)$ and $BDE(m^-—Na^+)$. For a very simple case, m = Cl, these are similar, 161 and 155 kcal mol^{-1}, respectively. The second aspect is attachment of the cation (H^+, Na^+) to the analyte. While proton affinities can be very large, with amides having proton affinities higher than 200 kcal mol^{-1}, alkali metal ion affinities are an order of magnitude smaller, frequently in the 20–30 kcal mol^{-1} range for polar molecules. Thus, the small alkali metal ion affinities make process (15) substantially endothermic and, if both processes (14) and (15) could be photoinduced, one would certainly expect them to have considerably different thresholds. This evaluation assumes that intact NaCl molecules exist in the solid MALDI target. This may not be so. The matrix may 'solvate' the Na^+ and Cl^- ions in the solid state, leading to a charge separation, so $BDE(Na^+—Cl^-)$ may not be relevant. Similar considerations were suggested by Vertes and co-workers[43] in their analysis of ionic derivatives of analytes and their improved performance in MALDI.

Mechanism 5 considers the possibility of the analyte ion having the form of a protonated salt molecule, not a sodiated acid molecule. This is an attractive mechanism because it is essentially the same as the mechanisms for protonation, with the analyte being in the form of the sodium salt. In mass spectrometry, we have some experience in processes that desorb large intact molecules, but the desorption of salts is not as common. While one may not expect to desorb intact salts in FAB, since the ionic components would be separated in the polar matrix, the case may be different for solid-state targets. Certainly, alkali metal halides are used when calibrating FAB instruments, because a variety of ionic species over a wide mass range are generated on bombardment of solid and mixed alkali metal halides. In MALDI, where the matrix-to-analyte ratio is high, isolated salts of analytes may be present in the solid target, which may desorb under conditions very similar to those responsible for analyte desorption in the acidic form. The discussion on whether a sodiated molecule is $\{M\}Na^+$ or $\{M - H + Na\}H^+$ has appeared when such ions are generated in FAB experiments. However, our results for methyl esters (Fig. 9), and others that have been recently reported,[19] suggest that the structure is a simple sodium ion adduct, negating this mechanism. However, the ions that contain two sodiums may in fact have salt-character.

Mechanism 6 proposes a gas-phase mechanism, which we know can occur under low-field conditions in the ion source, if desorbed molecules and alkali metal ions are present in the gas phase. This would be consistent with the data shown in Fig. 6. We have presented data which show that the formation of sodium ion adducts is matrix dependent, and tracks the intensity of the Na^+ peak. That is, in sinapinic acid, when NaCl is added, Na^+ adducts of analyte and matrix are not formed, and there is essentially no Na^+ peak. The reverse situation holds for 2,5-dihydroxybenzoic acid. This would be consistent with a gas-phase mechanism. However, Karas *et al.*[1] have reported an experiment with different results. When 2-nitrophenyl octyl ether (NPOE) is used as the matrix, and sodium salts are present, at threshold irradiance they detect only $[M + Na]^+$ ions—no ions from the matrix and no signal at M/z 23. At higher irradiances, the analyte signal becomes predominantly $[M + H]^+$. We note that the data in Fig. 1 are for low-power experiments, just above threshold. While sodiated adducts can dominate for 2,5-dihydroxybenzoic acid, this is a threshold observation—protonated analytes can be enhanced if higher powers are used. However, the fact that experiments can be done in which sodiated analytes can be formed without Na^+ ions being formed would negate a gas-phase mechanism. It depends in part on how different these experiments are, since the matrix that Karas *et al.* used (NPOE) in their MALDI experiment is a liquid.[1]

While the data reported contribute to the overall set of data in the literature, with which any proposed mechanism must be consistent, clearly other types of experiments must be done before the formation of sodiated species can be seriously addressed. Some of the same considerations cited in the discussion of protonated molecules should also be important. For example,

any mechanism should explain why one detects $[M + Na]^+$ under some conditions, but usually not $[M + nNa^+]$. If the process simply involves the desorption of $[M + Na]^+$ species that reside in the condensed phase, then higher concentrations of NaCl would seem to lead to multiply sodiated peptides (with multiple charges), which they do not.

Perhaps a very simple process, reported in the literature for improving MALDI resolution, provides the most powerful glimpse into the location of NaCl in the MALDI matrix. Beavis and Chait[44] discussed the fact that, if NaCl is present, alkali metal ion adducts can skew analyte peaks, making m/z determination difficult. When this is encountered, one can simply wash the MALDI probe with cold water, which eliminates much of the NaCl, as seen by a decrease in adduct ions. If the NaCl can be simply washed from the analyte–matrix solid, and spectral changes result, then the spectral changes cannot be due to sodium salts of the analyte deposited throughout the matrix, or dispersed, solvated Na^+ and Cl^- ions throughout the matrix. The effectiveness of the process shows fairly clearly that NaCl precipitates after the matrix to a large extent, and must be largely deposited on the outside of the crystals that are formed.

In a future publication, we shall evaluate aspects of the precipitation process which may provide a useful format in which to propose mechanisms. If one considers the concentrations and solubilities of sinapinic acid, a peptide, and NaCl in a typical solvent such as acetonitrile–water, then one may expect that on solvent evaporation the sinapinic acid would precipitate first, with the peptide precipitating next on top of the sinapinic acid crystals, and most soluble component, the salt, coming out of solution last. Such 'layering' could vary with component solubility and other aspects of the process through which MALDI targets are formed, which could be important in beginning to define the mechanism for sodium ion adduct formation and the apparent competition between sodiation and protonation processes in some matrices.

Some aspects of the correlation discussed between the proton affinity of analytes and the MALDI response may be related to the aspects of precipitation of species from solution. Recall that the data presented in Table 1 showed that analytes with proton affinities greater than 224 kcal mol^{-1} will be ionized in sinapinic acid, for example. However, we also showed that little or no response is seen for hexaglycine in sinapinic acid, and this peptide is one of the few for which a proton affinity is available, 236 kcal mol^{-1}. Thus, the data for small molecules in Table 1 suggest that a peak should have been detected for protonated hexaglycine. Does this negate the proton affinity correlation? It may show that proton affinity alone is insufficient to predict response. Aromatic molecules such as thymidine and phenazine are chemically similar to the matrices used here, also aromatic compounds. Thus, one might expect considerable analyte–matrix interactions during deposition of the MALDI target. This may be an additional requirement for the MALDI response. If the four small peptides used in the study shown in Fig. 4 are considered, the two that are most detectable, hexatyrosine and pentaphenylalanine, both have aromatic side-chains and are thus most similar, chemically, to the matrix. One possible explanation for the lack of response for hexaglycine and hexaalanine, is that, although they have sufficiently high proton affinities, they may not be incorporated into the matrix in a manner that is required for protonation and desorption.

CONCLUSIONS

In mass spectrometry, we have encountered a variety of new desorption/ionization techniques whose applications preceded our understanding of the mechanisms involved in the conversion of condensed-phase materials into gas-phase ions. In the evolution of these methods, models evolved as more complete descriptions of the chemical systems were realized. We are still trying to describe the chemical system of MALDI and, as the definition continues to evolve, so will the models. We must be cautious in how we define the chemical system. For example, one commonly cited characteristic is the tremendous sensitivity of MALDI. It works best with picomolar amounts of sample, in contrast to FAB in which nanomolar amounts are commonly required. However, one could also realize that ions are generated from the analyte for many minutes in FAB. In MALDI, one may obtain a few hundred spectra, with desorption/ionization taking place in 5–10 ns per laser pulse. Ions are only generated for a total period of approximately 1 μs. Thus, while MALDI requires three orders of magnitude less sample than FAB, FAB generates ion current for seven orders of magnitude longer time. The point is that the system must be carefully defined when a model is being pursued.

How are protonated analyte molecules formed in MALDI? For most of the experiments, a solid-state excited-state acid–base reaction seems reasonable. It allows for the formation of both protonated matrix and analyte molecules, and would also be consistent with cases where only protonated analytes are formed, in the absence of matrix ions. However, in most MALDI-MS experiments, matrix ions *are* formed. The model may change as we attempt to determine if those infrequent experiments, where analyte ions but not matrix ions are formed, are sufficiently common to warrant inclusion in the definition of the chemical system that we are trying to describe.

The same is true for alkali metal ion adduct formation. Were it not for a report of MALDI results in which, at threshold, $[M + Na]^+$ ions are formed for the analyte, with no other ionic species detected, then the chemical description would change. We believe that the description of $[M + Na]^+$ ions as protonated sodium salts is intriguing because it should have to be a contributing process to some extent. Protonation does occur and it is likely that salts of peptides are formed in the precipitate. In ternary mixtures involving matrix, analyte and NaCl, we clearly have too little information on the nature of the solid, dispersion of ionic components, etc. Indeed, we know little of the details of the solid composition when matrix and analyte alone are deposited. Perhaps the simplest model is that there is no

single model. Numerous mechanisms are operative and contribute to the observed spectrum. We certainly believe this to be the case when considering protonated matrix molecules and ionized matrix ions—they need not be related through a common mechanism.

Acknowledgements

Mass spectral data were acquired at the MSU Mass Spectrometry Facility, which is supported, in part, by a grant (RR-00480) from the Biotechnology Research Technology Program of the National Center for Research Resources of the NIH.

REFERENCES

1. M. Karas, D. Bachmann, U. Bahr and F. Hillenkamp, *Int. J. Mass Spectrom. Ion Processes* 78, 53 (1987).
2. M. Karas, U. Bahr and U. Giebmann, *Mass Spectrom. Rev.* 10, 335 (1991).
3. R. Cotter, *Anal. Chem.* 64, 1027A (1992).
4. B. Chait and S. Kent, *Science*, 257, 1885 (1992).
5. H. Egge, J. Katalinic, M. Karas and B. Stahl, *Pure Appl. Chem.* 63, 491 (1991).
6. P. Juhasz and C. Costello, *J. Am. Soc. Mass Spectrom.* 3, 785 (1992).
7. K. Tang, S. Allman and C. Chen, *Rapid Commun. Mass Spectrom.* 6, 365 (1992).
8. G. Parr, M. Fitzgerald and L. Smith, *Rapid Commun. Mass Spectrom.* 6, 369 (1992).
9. E. Nordhoff, A. Ingendof, R. Cramer, A. Overberg, B. Stahl, M. Karas, F. Hillenkamp and P. Crain, *Rapid Commun. Mass Spectrom.* 6, 771 (1992).
10. U. Bahr, A. Deppe, M. Karas and F. Hillenkamp, *Anal. Chem.* 64, 2866 (1992).
11. P. Danis, D. Karr, F. Mayer, A. Holle and C. Watson, *Org. Mass Spectrom.* 27, 843 (1992).
12. A. Vertes, R. Gijbels, R. D. Levine, *Rapid Commun. Mass Spectrom.* 4, 228 (1990).
13. D. R. Knapp, *Methods Enzymol.* 193, 314 (1990).
14. K. L. Light, D. B. Kassel and J. Allison, *Biomed. Environ. Mass Spectrom.* 18, 18 (1989).
15. M. Medina, T. Huth-Fehre, A. Westman and B. U. R. Sundqvist, *Org. Mass Spectrom.* 29, 207 (1994).
16. See, for example: T-W. D. Chan, A. W. Colburn, P. J. Derrick, D. J. Gardiner and M. Bowden, *Org. Mass Spectrom.* 27, 188 (1992).
17. K. Zhang, D. M. Zimmerman, A. Chung-Phillips and C. J. Cassady, *J. Am. Chem. Soc.* 115, 10812 (1993).
18. Z. Wu and C. Fenselau, *Rapid Commun. Mass Spectrom.* 6, 403 (1992).
19. B. Wang, K. Dreisewerd, U. Bahr, K. Karas and F. Hillenkamp, *J. Am. Soc. Mass Spectrom.* 4, 393 (1993).
20. R. C. Beavis, *Org. Mass Spectrom.* 27, 653 (1992).
21. M. Gimon, L. Preston, T. Solouki, M. White and D. Russell, *Org. Mass Spectrom.* 27, 827 (1992).
22. H. H. Jaffe and H. L. Jones, *J. Org. Chem.* 30, 964 (1965).
23. R. C. Beavis, T. Chaudhary and B. Chait, *Org. Mass Spectrom.* 27, 156 (1992).
24. C. D. Parker and D. M. Hercules, *Anal. Chem.* 58, 25 (1986).
25. S. G. Lias, J. E. Bartmess, J. F. Liebman, J. L. Holmes, R. D. Levin and W. G. Mallard, *J. Phys. Chem. Ref. Data* 17, Suppl. 1 (1988).
26. H. Ehring, M. Karas and F. Hillenkamp, *Org. Mass Spectrom.* 27, 472 (1992).
27. C. D. Mowry and M. V. Johnston, *Rapid Commun. Mass Spectrom.* 7, 569 (1993).
28. M. Karas, *Analusis* 20, 31s (1992).
29. R. C. Beavis and B. T. Chait, in *Methods and Mechanisms for Producing Ions from Large Molecules*, edited by K. G. Standing and W. Ens, p. 227, Plenum Press, New York (1991).
30. J. C. Rouse and J. Allison, *J. Am. Soc. Mass Spectrom.* 4, 259 (1993).
31. T. Cornish and R. J. Cotter, *Rapid Commun. Mass Spectrom.* 6, 242 (1992).
32. J. A. Castoro, C. Koster and C. L. Wilkins, *Anal. Chem.* 65, 784 (1993).
33. J. A. Castoro and C. L. Wilkins, *Anal. Chem.* 65, 2621 (1993).
34. J. A. Castoro, C. Koster and C. L. Wilkins, *Rapid Commun. Mass Spectrom.* 6, 239 (1992).
35. R. Hettich and M. Buchanan, *J. Am. Soc. Mass Spectrom.* 2, 402 (1991).
36. J. C. Dunphy, K. L. Busch, R. L. Hettich and M. V. Buchanan, *Anal. Chem.* 65, 1329 (1993).
37. E. A. Stemmler, R. L. Hettich, G. B. Hurst and M. V. Buchanan, *Rapid Commun. Mass Spectrom.* 7, 828 (1993)
38. W. Ens, Y. Mao, F. Mayer and K. G. Standing, *Rapid Commun. Mass Spectrom.* 5, 117 (1991).
39. D. Bombick, J. D. Pinkston and J. Allison, *Anal. Chem.* 56, 396 (1984).
40. P. Juhasz and C. E. Costello, *Rapid Commun. Mass Spectrom.* 7, 343 (1993).
41. H. Wu and J. Allison, *J. Am. Soc. Mass Spectrom.* 5, 564 (1994).
42. D. B. Kassel and J. Allison, *Biomed Environ. Mass Spectrom.* 7, 221 (1988).
43. J. Claereboudt, M. Claeys, H. Geise, R. Gijbels, and A. Vertes, *J. Am. Soc. Mass Spectrom.* 4, 798 (1993).
44. R. C. Beavis and B. T. Chait, *Anal. Chem.* 62, 1836 (1990).

APPENDIX

Proton affinities, ionization energies and hydrogen atom affinities of the compounds listed in Table 1

The proton affinities of all the compounds in Table 1 are known.[A1] The ionization energies of phenazine, acridine and 1,8-bis(dimethylamino)naphthalene were obtained from the data compiled by Lias *et al.*[25] The ionization energies of 4-cyanobenzaldehyde, thymidine and 1,7-diaminoheptane were not available in the compilation. The ionization energy of 4-cyanobenzaldehyde was estimated by considering the ionization energies of similar compounds, fluorobenzene (IE = 9.20 eV), 1-fluoro-4-cyanobenzene (IE = 9.74 eV), nitrobenzene (IE = 9.86 eV), 1-nitro-4-cyanobenzene (IE = 10.23 eV), aniline (IE = 7.72 eV) and 4-cyanoaniline (IE = 8.17 eV).[25] These data suggested that the cyano group in the *para*-position would raise the ionization energy of a compound by]0.45 eV. Since the ionization energy of benzaldehyde was reported to be 9.49 eV,[25] the ionization energy of 4-cyanobenzaldehyde was thus estimated as 9.94 eV, or 229 kcal mol^{-1}.

The ionization energy of thymidine was also estimated. The ionization of thymine was reported to be 8.8 eV.[25] The ionization energies of acetamide and *N*-methylacetamide were reported to be 9.65 and 9.3 eV respectively,[25] suggesting that the *N*-methyl derivative of the amide group would lower the ionization energies by 0.3 eV. The ionization energies of compounds with cyclic ether or hydroxyl group were all above 9.2 eV tetrahydrofuran (IE = 9.41 eV), *n*-butanol

ΔE = 10
ethoxyet
desoxym
ionizatio
tion en
8δ – 0.

The
also esti
butane
and 10
dichloro
be 10.55
adding
opposite
ionizatio
energy o
of the io
tunately.

(IE = 10.06eV), cyclobutanol (IE = 9.25 eV) and 2-ethoxyethanol (IE = 9.6 eV),[25] suggesting that the 2-desoxyriboside group of thymidine have a higher ionization energy than the thymine portion. The ionization energy of thymidine was thus estimated to be 8.8 − 0.3 = 8.5 eV, or 196 kcal mol^{-1}.

The ionization energy of 1,7-diaminoheptane was also estimated. The ionization energies of 1,4-dibromobutane and 1-bromobutane were reported to be 10.15 and 10.13 eV, respectively,[25] and the those of 1,3-dichloropropane and 1-chloropropane were reported to be 10.85 and 10.82 eV, respectively,[25] suggesting that adding the second identical functional group on the opposite end of an alkyl chain would not change the ionization energy significantly. Therefore, the ionization energy of 1-aminoheptane would be a good estimation of the ionization energy of 1,7-diaminoheptane. Unfortunately, the ionization energy of 1-aminoheptane is not available. However, the ionization energies of smaller aminoalkanes are available: 1-aminopropane (IE = 8.97 eV), 1-aminobutane (IE = 8.71 eV), 1-aminopentane (IE = 8.67 eV), 1-aminohexane (IE = 8.63 eV) and 1-aminooctane (IE = 8.5 eV).[25] The ionization energy of 1-aminoheptane was estimated to be 8.6 eV, or 198 kcal mol^{-1}. This is also used as the ionization energy for 1,7-diaminoheptane.

The hydrogen atom affinity of a molecular ion, $M^{+\cdot}$, was defined as

$$[M + H]^{+} \rightarrow M^{+\cdot} + H^{\cdot}$$

ΔH = hydrogen atom affinity, $HAA(M^{+\cdot})$

Also, $HAA(M^{+\cdot}) = IE(M) + PA(M) - IE(H)$. Using the ionization energies and proton affinities listed in Table 1, and $IE(H)$ = 313.7 kcal mol^{-1}, the hydrogen atom affinities were estimated.

REFERENCES

A1. S. G. Lias, J. F. Liebman and R. D. Levin, *J. Phys. Chem. Ref. Data* 13, 695 (1984).

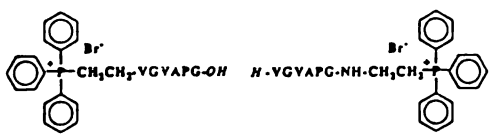
Appendix II

JOURNAL OF MASS SPECTROMETRY, VOL. 30, 000–000 (1995)

Dear Sir,

Enhanced Detection of Peptides in Matrix-assisted Laser Desorption/Ionization Mass Spectrometry Through the Use of Charge-localized Derivatives

In fast atom bombardment (FAB) mass spectrometry, charged components of ionic analytes are referred to as 'preformed ions' in the liquid FAB target. These include inorganic ions such as Na^+ and organic cations such as tetraalkylammonium ions. When cationic forms of the analyte predominate in the FAB matrix, strong signals are usually observed. This suggests that, at least for positive ions, such ionic species are desorbed directly.[1] In contrast, some analytes are desorbed in neutral form and are subsequently ionized in the gas phase by ion–molecular reactions.[2,3] Thus, preformed ions yield an enhanced response because these analytes need only be desorbed; no ionization step is required. Such concepts have led to the development of charge-localized derivatives for peptides[4] as a chemical tool to use in FAB analyses. For example, derivatization procedures have been reported that will place a charged triphenylphosphium (TPP) group on either the *C*- or *N*-terminus of a peptide.[5] The actual group appended to the *N*-terminus is the ethyltriphenylphosphonium group, with the aminoethyltriphenylphosphonium group added to the *C*-terminus. The Br^- salts of two such derivatives of the hexapeptide *H*-VGVAPG-*OH*, are shown in structures I and II. Two advantages of TPP derivatization of peptides in FAB analyses[6] have been discussed. Detection limits are lowered, and sequence information is contained in a smaller number of fragment ion peaks. While protonated peptides yield both *N*- and *C*-terminal fragment ions, *C*-terminal TPP derivatives yield only *C*-terminal ions.[6]



I II

Although there may be aspects of TPP-derivatized peptides related to their behavior in FAB that are not fully understood, such as some surface activity in polar solvents, the preformed ion concept does appear to be useful; it is leading to the development of new analytical methods using FAB.

This letter reports that TPP-derivatized peptides also exhibit an enhanced response in matrix-assisted laser desorption/ionization (MALDI).[7] This is not an obvious result, since the relationship between the chemical details of the solid-state MALDI target and resulting gas-phase ion abundances is not understood. In particular, we shall use as an example a relatively small hexapeptide. The peptide *H*-VGVAPG-*OH* was selected because it yields a very weak signal representing the protonated molecule in positive-ion MALDI. Higher powers are required to generate ions from this small peptide than are needed to generate gas-phase protonated insulin, for example. At higher powers, mass spectral peaks broaden and the resolving power is degraded.

In Fig. 1, the MALDI spectrum of this peptide, and that for the *N*- and *C*-terminal TPP-derivatized peptides, prepared according to published procedures[5] and isolated as the Br^- salts, are shown. All three spectra were obtained at the same laser power; α-cyano-4-hydroxycinnamic acid was used as the matrix. The data were obtained on a Model VT2000 time-of-flight MALDI instrument (Vestec, Houston, TX, USA). This is a 30 kV instrument, which utilizes a nitrogen laser and a transient recorder which acquires spectra with 5 ns resolution. For the data shown, 1 μl of an analyte solution and 1 μl of the saturated matrix solution (2:1 (v/v) water–acetonitrile solvent) were applied to the probe tip, and allowed to air dry. Each spectrum shown represents the sum of data from 64 laser shots. In making TPP-derivatized peptides, the chemistry was performed on 10 nmol samples of the peptide. The solution containing the derivatized product was then diluted to a concentration of 1 pmol μl^{-1} if the derivatization reactions were quantitative, which they are not. Thus, the spectrum shown in Fig. 1(a) results from 1 pmol of peptide on the probe, with less than 1 pmol of the derivatized peptides being analyzed in the data shown in Fig. 1(b) and (c).

For the conditions chosen, there is no detectable signal for the peptide before derivatization. However, very strong signals were observed for the TPP-derivatized peptide. Such signal enhancements were also observed when sinapinic acid or 2,5-dihydroxybenzoic acid was used as the matrix. When sinapinic acid is used as the matrix, a peak representing Na^+ is always observed, unless the matrix is purified. The TPP-derivatized peptide yields a strong signal at laser powers even lower than those required to yield signals for Na^+. For the experiment represented by Fig. 1(a), even when the laser power is doubled no signal representing the underivatized peptide is detected. Hence ionic species in the MALDI experiment can behave as they do in FAB, leading to enhanced sensitivity when only a desorption step is required.

In addition to the obvious enhancement in detectability realized when the TPP derivatives are made, we suggest that such molecules may play an important role in investigating the mechanisms through which ions are generated in MALDI. Why are so few compounds suitable as MALDI matrices? This is a difficult question to pursue, since multiple processes, both desorption and ionization, must occur. TPP-derivatized peptides may be useful chemical probes since the charge-localized peptide ions can be desorbed directly; no proton transfer is required. This may allow for independent studies of the desorption and ionization steps.

This research was performed in the MSU Mass Spectrometry Facility, which is suppported, in part, by a grant (RR-0480) from the Biotechnology Research Technology Program of the National Center for Research Resources of the NIH.

Yours

PAO-CHI LIAO and JOHN ALLISON*
Department of Chemistry,
Michigan State University,
East Lansing,
Michigan 48824-1322, USA

CCC 1076–5174/95/000000–00

Received 8 November 1994
Accepted 14 December 1994

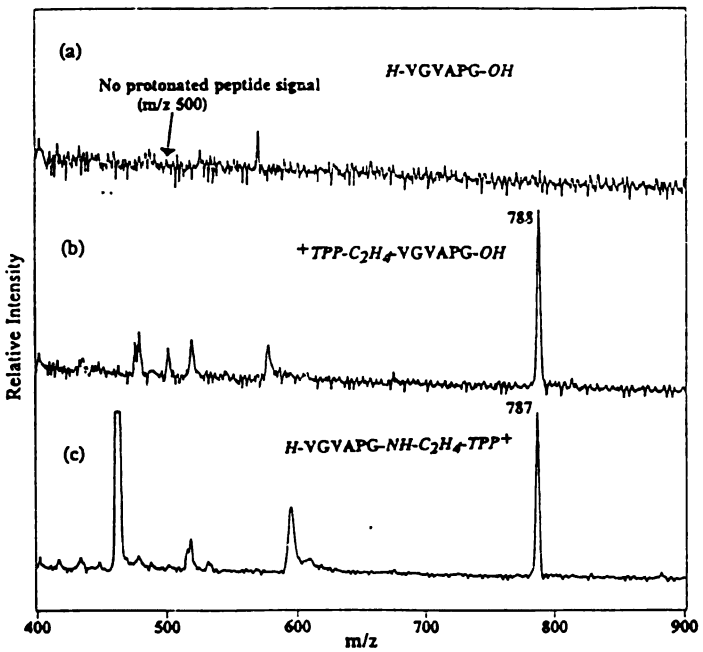


Figure 1. Portions of the MALDI mass spectra for the peptide *H*-VGVAPG-*OH* (molecular mass 498.6). (a) Underivatized peptide; (b) *N*-terminal ethyltriphenylphosphonium derivative; (c) *C*-terminal aminoethyltriphenylphosphonium derivative. The same matrix and the same laser power were used to obtain each spectrum.

References

1. R. G. Cooks and K. L. Busch, *Int. J. Mass Spectrom. Ion Phys.* 53, 111 (1983).
2. J. Sunner, A. Morales and P. Kebarle, *Int. J. Mass Spectrom. Ion Processes* 86, 169 (1988).
3. J. C. Rouse and J. Allison, *J. Am. Soc. Mass Spectrom.* 4, 259 (1993).
4. See, for example, J. E. Vath and K. Biemann, *Int. J. Mass Spectrom. Ion Processes* 100, 287 (1990).
5. D. S. Wagner, A. Salari, D. A. Gage, J. Leykam, J. Fetter, R. Hollingsworth and J. T. Watson, *Biol. Mass Spectrom.* 20, 419 (1991).
6. J. T. Watson, D. S. Wagner, Y.-S. Chang, J. R. Strahler, S. M. Hanash and D. A. Gage, *Int. J. Mass Spectrom. Ion Processes* 111, 191 (1991).
7. P.-C. Liao and J. Allison, in *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics, San Francisco, CA, May 30–June 4, 1993.*

Appendix III

JOURNAL OF MASS SPECTROMETRY, VOL. 30, 000–000 (1995)

JMS Letters

Dear Sir,

Dissecting Matrix-assisted Laser Desorption/Ionization Mass Spectra

Matrix-assisted laser desorption/ionization (MALDI)[1] mass spectrometry (MS) is a technique that is evolving at rapid rate. Readers who have been dazzled by the detection limit and molecular mass barriers of MS that MALDI has shattered may have a very different view of the method from users of MALDI MS. When faced with an electron impact ionization mass spectrum of a volatile organic compound, most mass spectrometrists understand the correct context in which to interpret the spectrum, that is, they have a reasonable idea of the experimental conditions and how the spectrum was probably obtained. This is not the case with MALDI MS. Although many MALDI mass spectra have been published, there is no well defined entity that we can refer to as *the* MALDI mass spectrum of a compound. This is true for other desorption/ionization techniques also, such as fast atom bombardment.[2]

Figure 1 shows a MALDI mass spectrum of a mixture of bradykinin (B) and bovine insulin (I), using the matrix α-cyano-4-hydroxycinnamic acid (M). The data were obtained on a Vestec (Houston, TX, USA) ResearchTec time-of-flight MALDI instrument. This instrument employs an accelerating voltage in the ion source of up to 30 kV; it utilizes a nitrogen laser and a transient recorder which acquires spectra with 5 ns resolution. For the data shown, 1 μl of an analyte solution and 1 μl of the saturated matrix solution (water–acetonitrile (2:1, v/v)) were applied to the probe tip and allowed to air dry, leaving 18 nmol of matrix, 2 pmol of B and 2 pmol of I on the 3.5 mm^2 target. This letter discusses how such MALDI spectra evolve and approaches that can be used to improve mass spectral quality.

Obtaining a MALDI mass spectrum is, at least on some instruments, like playing a video game; 'joysticks' are frequently an integral part of the hardware. The pulsed laser is tightly focused on a small number of crystals of the target (several research groups are investigating alternative approaches to MALDI target preparation so that it is not necessary to search for 'good spots' on the probe tip; see, e.g., Ref. 3). Data acquisition frequently involves a digital oscilloscope which sums or averages some number of successive transients. The laser firing begins, and the operator watches a display as spectra continue to accumulate. With peaks cease growing with continuing laser pulses, it is time to move the laser spot quickly to another location on the probe or adjust the laser irradiance, with the hope that only a small number of spectra which contain no information will be added into the accumulating data before a new region is found which will generate additional analyte ions. Frequently, hundreds of transients, of varying quality, are accumulated to yield a single MALDI spectrum.

We have found that the experiment is much easier do when a much larger spot size is used (and the laser power is concomitantly to re-establish the laser fluence required for the MALDI process). In Big Spot MALDI, instead of constantly

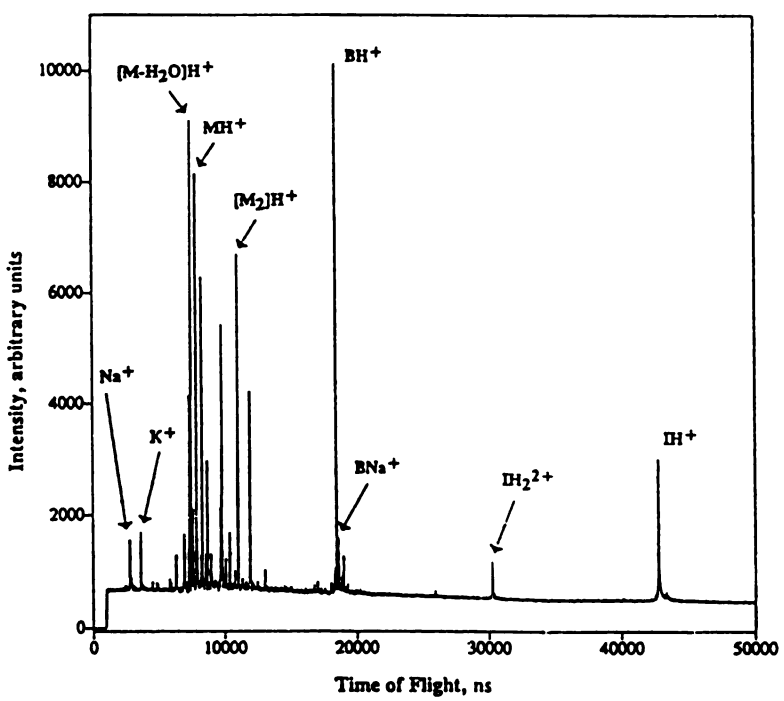


Figure 1. MALDI time-of-flight mass spectrum of 2 pmol of bradykinin (B) and 2 pmol of insulin (I), using the matrix α-cyano-4-hydroxycinnamic acid (M). This spectrum represents the average of 100 transients (100 pulsed laser irradiations of the sample) using Big Spot MALDI. Peaks range from *m/z* 23 (Na^+) to 5734.6 (IH^+).

hunting for single crystals that will yield ions when irradiated, many such crystals are irradiated simultaneously. We frequently irradiate 20–30% of the target area with the laser. This improves the reproducibility and allows us to begin actually to make comparisons between successive experiments.

Figure 1 represents the sum of 100 time-of-flight transients. With it, we can begin to make some observations about the MALDI experiment. It appears that Na^+ is formed, and there is a small, resolved peak just above that for BH^+ which represents BNa^+. Perhaps this suggests that, when Na^+ adducts are formed, a mechanism for forming free Na^+ must also exist.[4] One may consider the relative intensities of the I- and B-related peaks as providing some information on the relative responses of MALDI for these two compounds. Resolution for the experiment can also be calculated based on the spectrum provided.

To generate this spectrum, the experiment was configured so that laser fires 10 times in 2.5 s. The laser then stops while the transient recorder downloads the file representing the average of these ten, and a new file representing the next 10 spectra is generated. Figure 2 shows ten successive spectra obtained in this way, each an average of 10 transients. These 10 averaged spectra were then averaged to yield *the* MALDI spectrum in Fig. 1. In collecting these data, the laser spot remained in a single position. (For this demonstration, a set of averaged spectra were manipulated, instead of a larger collection of single transients. This was only done so the set of spectra to be considered could be viewed with a reasonable signal-to-noise ratio.)

Much can be learned of the MALDI process using Big Spot MALDI, with the spectrum dissected as shown in Fig. 2. In spectrum 1, there are substantial peaks representing Na^+, BH^+, BNa^+ and IH^+. After 40 shots, spectra 5 and 6 contain no information. Then peaks reappear, which again are gone after several tens of laser shots.

The MALDI data obtained and presented in this form provide much more information on the experiment and will lead to higher quality spectra. Clearly, if one desires to obtain a single spectrum from the experiment, one should not include all of the spectra of Fig. 2 in the composite, since at least three of the spectra contain only noise. A few other observations are of note. The Na^+ peak decreases with time much more quickly than does the peak representing MH^+. Also, the peak representing BNa^+ decreases with time more quickly than does that representing BH^+ Hence, in this single data set, one can begin to identify correlations between Na^+ and Na^+ adducts of analytes. While the averaged spectrum contains peaks representing Na^+ and BNa^+, all of the spectra accumulated do *not* contain these ions, some only contain peaks representing protonated molecules. Also, the IH^+ peak intensity decreases more rapidly than that for BH^+ throughout the experiment.

Given a *set* of spectra instead of one spectrum, one may choose to use single spectra from the set shown in Fig. 2 for extracting information. For example, the resolution based on analysis of the IH^+ peak changes from 365 (spectrum 1) to 478 (spectrum 4). This makes sense, since the IH^+ peak probably contains an unresolved contribution due to INa^+ in spectrum 1 which is not significantly contributing by spectrum 4.

With this approach to data acquisition, and using Big Spot MALDI, one can begin to appreciate how averaged spectra evolve, and what must be avoided when spectra of the highest quality are desired. Figure 3(a) and (b) show expansions of two peaks from Fig. 1 representing MH^+ and BH^+, respectively. Data from all 10 of the averaged spectra in Fig. 2 for

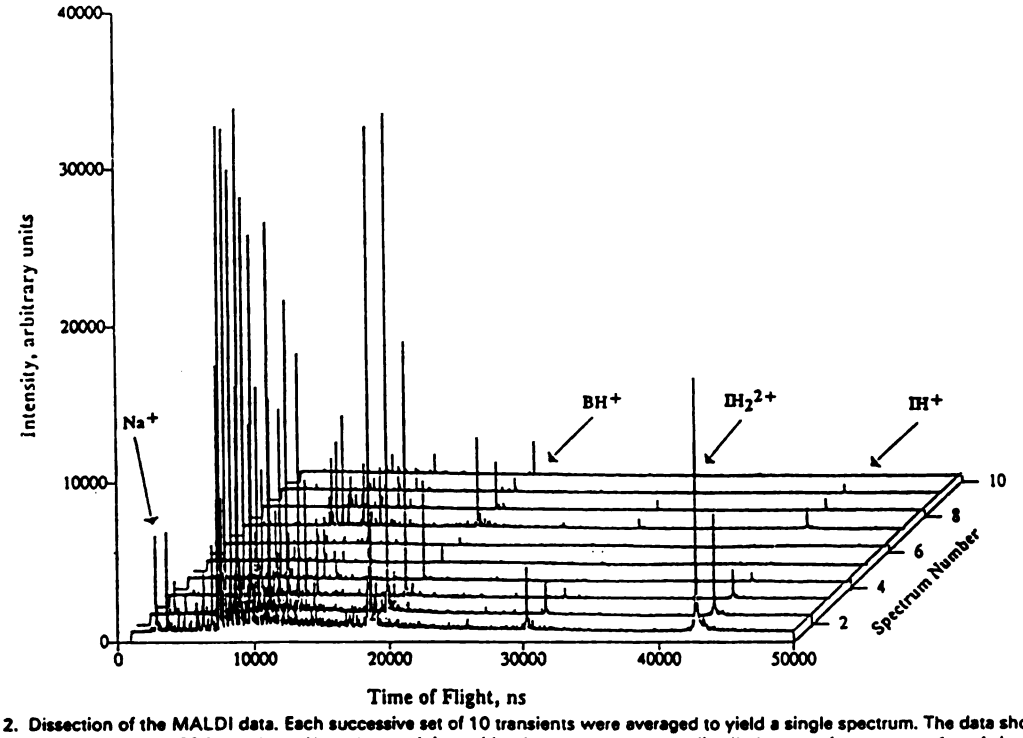


Figure 2. Dissection of the MALDI data. Each successive set of 10 transients were averaged to yield a single spectrum. The data shown, 10 averaged spectra, represent 100 laser shots. Note that peak intensities (raw data, not normalized) decrease from spectra 1 to 4. Ion generation commences again after an additional 20 laser shots.

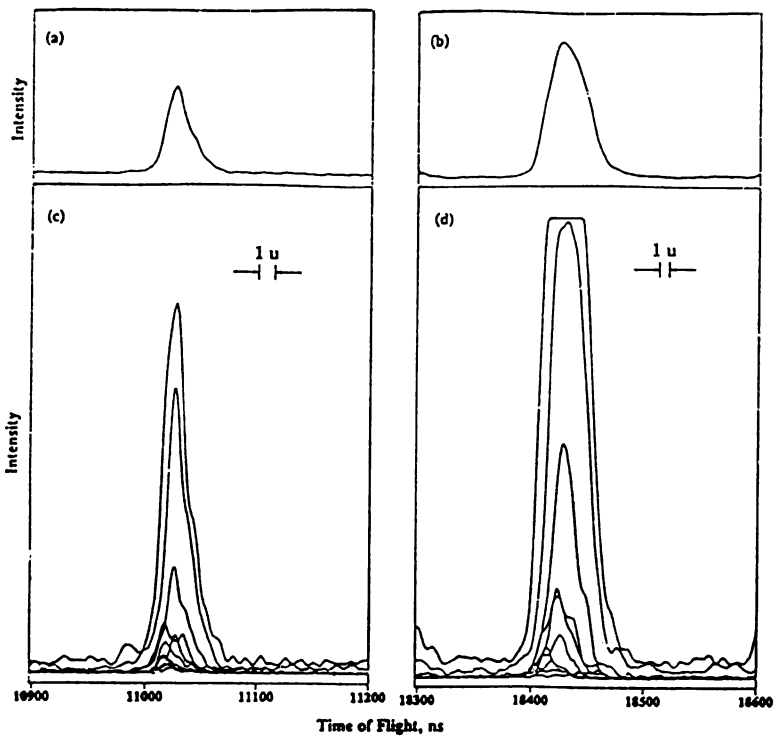


Figure 3. Parts (a) and (b) show expanded views of two peaks from Fig. 1, representing MH^+ and BH^+ respectively. Parts (c) and (d) show the same regions from all 10 spectra of Fig. 2 that were averaged to yield Fig. 1. The approximate unit of flight time representing 1 u for each region of the transient are also indicated in (c) and (d). Clearly, resolution of single traces in (d) is higher than in the composite (b).

these regions of the time-of-flight spectra are shown in Fig. 3(c) and (d). In comparing Fig. 3(a) and (c), it is clear that a collection of very different peak shapes come together to yield the average. Figure 3(d) shows what can also be encountered when spectra of varying quality and intensity are averaged; signals are sufficiently intense in one spectrum such that a peak is saturated. Clearly, such contributions must be avoided, since including a flat-topped peak in an average can only lead to peak distortions, lower resolution and less accurate peak centroiding capabilities. The contribution of higher mass components to the overall peak shape in Fig. 3(b) is certainly clear from the data as presented in Fig. 3(d). Thus, some spectra in the set in Fig. 2 should be deleted before an overall average mass spectrum for the experiment is generated, because they contain saturated peaks. This would be important if the peak in question represented an unknown whose molecular mass was desired, to the highest accuracy that the experiment would allow.

Why do the spectra change with time? We suggest that we are showing, in some crude way, depth profiling of the MALDI target in Fig. 2. Sodium and potassium salts, the most soluble components of the solution deposited on the probe, precipitate last, hence signals representing Na^+ and K^+ are present in the early spectra and are quickly lost. The data also suggest that while I and B coprecipitate, B appears to come out of solution first. The fact that ion generation ceases after 40 laser pulses but begins again later could have many causes which cannot be evaluated here. This may be purely physical in nature, with fresh crystals falling from the sides of the laser ablation region into the irradiated area. We do note, however, that when ion generation ceases (after spectrum 4 in Fig. 2), it can be stimulated again in the same location by changing the power to a new, higher value.

Obviously, with a collection of time-dependent spectra such as those shown in Fig. 2, one may quickly gain many insights into the MALDI process by generating mass chromatograms and total ion current profiles, as is done in gas chromatography/mass spectrometry.[5] This exercise has been very revealing in helping us to understand what we can do with MALDI data to optimize our analytical capabilities. Such time-dependent data acquisition may be an important probe of relative responses in mixture analyses by MALDI. Hopefully, the dissected MALDI spectrum presented here will also help those not using the technique to appreciate what goes into the single MALDI spectra that have appeared in the literature to date, which represent other than single-shot transients.

This research was performed in the MSU Mass Spectrometry Facility which is supported, in part, by a grant (RR-0480) from the Biotechnology Research Technology Program of the National Center for Research Resources of the NIH. The authors also acknowledge Dr Joseph Hiller, Ms Gabriella Székely and Mr Gary Levine for helpful discussions.

Yours

PAO-CHI LIAO and JOHN ALLISON†
Department of Chemistry,
Michigan State University,
East Lansing,
Michigan 48824-1322, USA

† Author to whom correspondence should be addressed.

References
1. F. Hillenkamp, M. Karas, R. C. Beavis and B. T. Chait, *Anal. Chem.* 63, 1193A (1991).
2. E.g. S. W. Lemire, X. Zhao, L. M. Tolbert and K. L. Busch, *J. Am. Soc. Mass. Spectrom.* 5, 1017 (1994).
3. O. Vorm, P. Roepstorff and M. Mann, *Anal. Chem.* 66, 3281 (1994).
4. P.-C. Liao and J. Allison, *J. Mass Spectrom. in press.*
5. J. T. Watson, *Introduction to Mass Spectrometry*, 2nd ed. Raven Press, New York (1985).

Appendix IV

ANALYTICAL BIOCHEMISTRY 219, 9–20 (1994)

An Approach to Locate Phosphorylation Sites in a Phosphoprotein: Mass Mapping by Combining Specific Enzymatic Degradation with Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry

Pao-Chi Liao,* Joe Leykam,† Philip C. Andrews,‡ Douglas A. Gage,† and John Allison*,[1]

*Departments of *Chemistry and †Biochemistry, Michigan State University, East Lansing, Michigan 48824; and ‡Department of Biochemistry, University of Michigan, Ann Arbor, Michigan 48109*

Received November 5, 1993

A rapid, picomole-scale method is described to locate phosphorylation sites in phosphoproteins by using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) combined with enzymatic modification of the analyte. There are three steps to locate phosphorylation sites in a phosphoprotein: (i) degradation of the phosphoprotein into small peptides by specific enzymatic or chemical reactions; (ii) identification of the phosphopeptides by −80 (or multiples of −80)-Da mass shifts in the mass spectra after dephosphorylation with alkaline phosphatase; (iii) location of the phosphorylation sites by mass mapping. As the size of the protein increases, it is advantageous to fractionate the mixture by HPLC and analyze each fraction by MALDI-TOF-MS. To perform mass mapping, the primary structure of the protein must be known. Bovine β-casein was analyzed by this method. The conclusions about the specific phosphorylation sites of bovine β-casein from our data coincide with previously reported results. From calculations, it is found that a mass spectrometer with 0.1% mass accuracy is sufficient, for mass mapping, to identify completely or partially digested tryptic peptides in the mass range of 100~8000 Da from bovine β-casein (MW 23,983).

Phosphorylation of serine, threonine, and/or tyrosine residues is a common post-translational modification of proteins. In eukaryotes, protein phosphorylation–dephosphorylation is one of the most important mechanisms for regulation of many intracellular functions (1). The growing number of publications concerning the phosphorylation sites of peptides and proteins shows that the field is very active.

Several methods have been developed and used to locate phosphorylation sites in phosphoproteins. All employ the strategy of degrading the phosphoprotein chemically or enzymatically into small peptides in the initial step, followed by analysis of the composition and sequence of each fragment. Phosphopeptide mapping has a long history, and has employed sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE)[2] (2), reverse-phase HPLC (3), or two-dimensional separation on TLC plates (4). However, peptide mapping experiments are tedious and often complicated by problems that arise during sample preparation, leading to inconclusive results (4). Identification of phosphorylation sites in phosphopeptides using automated Edman sequence analysis has also been accomplished (5–18). Phosphoserine is the most abundant, naturally occurring phosphorylated amino acid residue. Unfortunately, the phenylthiohydantoin (PTH) derivative of phosphoserine breaks down to free phosphate and PTH-dehy-

[1] To whom correspondence should be addressed at Department of Chemistry, Michigan State University, East Lansing, MI 48824. Fax: (517) 353-1793.

[2] Abbreviations used: MALDI-TOF-MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; PTH, phenylthiohydantoin; FAB-MS, fast atom bombardment mass spectrometry; LSIMS, liquid secondary ion mass spectrometry; TFA, trifluoroacetic acid; FMOC, 9-fluorenylmethoxycarbonyl For peptides with n residues $R_1R_2R_3R_4R_5 \ldots R_n$, ppp-$R_1R_2R_3R_4R_5 \ldots R_n$ is the designation used to indicate three phosphorylation sites. If they are specifically known, they are indicated as $R_1pR_2pR_3R_4pR_5 \ldots R_n$, with a lowercase p preceding the phosphorylated residue designation.

0003-2697/94 $6.00

droalanine in Edman degradation cycles and yields no identifiable eluting peak representing a phosphorylated fragment (5,6). In contrast, the PTH derivative of phosphotyrosine gives an identifiable peak on HPLC analysis (7). Two modified sequencing techniques have been proposed to solve this problem. Meyer and his coworkers chemically modified phosphoserine in the intact peptide prior to sequence analysis (5,6,8,9). Their method, which converts phosphoserine into detectable *S*-ethylcysteine, has been used successfully to locate phosphorylation sites in several phosphoproteins by microsequencing (10–15). This chemical modification approach, however, is not applicable to phosphorylated residues located at the C- or N-terminus, because the intermediates formed during the reaction are unstable (9). Another modified sequencing technique (16–18), which applies to not only phosphoserine but phosphotyrosine and phosphothreonine, involves the detection of the release of [^{32}P]P_i or other phosphorylated degradation products at the position of the phosphorylated residue during Edman degradation. However, the detection of ^{32}P radioactivity is the only criterion for localizing the phosphorylated amino acid.

Fast atom bombardment mass spectrometry (FAB-MS) and liquid secondary ion mass spectrometry (LSIMS) have been used as effective techniques to study the phosphorylation states of phosphopeptides (14,19–29). Two major advantages of mass spectrometric methods are the speed of analysis and the absence of a requirement for radioactive isotope labeling (27). Sample size requirements, however, are typically at the nanomole level for FAB-MS or LSIMS. This often exceeds the amount of material available, practically, from biological samples. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS), developed by Karas and Hillenkamp (30), has now been accepted as an extremely sensitive technique for the analysis of peptides and proteins (31,32). However, most MALDI-MS applications have focused on protein molecular weight determinations. Recently, MALDI and electrospray, both new ionization techniques, have been considered as MS tools for characterizing phosphopeptides (33–36). Results from both methods are frequently complementary. We present here a fast, picomole-scale methodology to locate the phosphorylation sites in phosphoproteins by taking advantage of the speed, high sensitivity, and mixture analysis capability of MALDI-TOF-MS, and the selectivity of specific enzymatic and chemical reactions. The phosphorylation sites of the protein Op18, in transformed lymphocytes, had been investigated by this method in our laboratory (37). The advantages, in comparison to other techniques, practical considerations, and current limitations of the proposed method are the subject of this work.

MATERIALS AND METHODS

Chemicals. β-Casein (from bovine milk), TRIZMA hydrochloride (reagent grade), and EDTA were purchased from Sigma Chemical Co. (St. Louis, MO). α-Cyano-4-hydroxycinnamic acid was purchased from Aldrich Chemical Co. (Milwaukee, WI). Acetonitrile and trifluoroacetic acid (TFA) were purchased from EM Science (Gibbstown, NJ). Ammonium bicarbonate was purchased from Mallinckrodt Specialty Chemical Co. (Paris, KY). Trypsin (sequencing grade, from bovine pancreas), endoproteinase Glu-C (sequencing grade, protease V8 from Staphylococcus aureus), alkaline phosphatase (from calf intestine), and guanidine hydrochloride were purchased from Boehringer-Mannheim Biochemicals (Indianapolis, IN). For HPLC solvents, acetonitrile was purchased from Burdick & Jackson, Baxter Healthcare Co. (Muskegon, MI), and TFA was purchased from Pierce Chemical Co. (Rockford, IL). Cation exchange resin (AG 50W-X8, mesh size 100–200) was purchased from Bio-Rad Laboratories, Inc. (Melville, NY).

Synthesis of model phosphopeptides. Phosphopeptides were synthesized by the University of Michigan Protein and Carbohydrate Structure Facility using FMOC/O*t*Bu methods. Phosphorylation of the nascent peptides was carried out on the resin using the global phosphorylation approach as described by D. M. Andrews *et al.* (38).

Enzymatic digestion. (i) Trypsin: 580 μg of bovine β-casein was mixed with 20 μg of trypsin in 200 mM Tris buffer (pH 8.0), 6 M guanidine, and 0.02% EDTA to give the substrate/enzyme ratio of ~30 (w/w). The mixture was incubated at 37°C for 24 h. For tryptic digestion of phosphopeptides, the same incubation conditions and substrate/enzyme ratio were used, but the amounts of substrate and enzyme were reduced. (ii) Alkaline phosphatase: A small amount of phosphopeptide (1 ~ 10 pmol) was mixed with 2 ~ 20 units of calf intestine alkaline phosphatase in 25 mM NH_4HCO_3 buffer (pH 8.0). The mixture was incubated at 37°C for 15 min to 4 h for dephosphorylation, or at room temperature for 6 h for on-probe dephosphorylation experiments. (iii) Endoproteinase Glu-C: 0.3 μg of the peptide of interest was mixed with 0.015 μg endoproteinase Glu-C in 20 μl of 25 mM NH_4HCO_3 buffer (pH 8.0) to give the substrate/enzyme ratio of ~20 (w/w). The mixture was incubated at room temperature for 4 h.

HPLC fractionation. The tryptic digest mixture (10 μg, ~400 pmol) of bovine β-casein was fractionated by reverse-phase HPLC with a 0–90% linear gradient of acetonitrile in 0.1% TFA and a flow rate of 50 μl/min on an Aquapore RP-300, 7 micron, 50 × 1.0-mm column (Applied Biosystems). Fifteen fractions were collected manually, each fraction corresponded to a resolvable peak in the chromatogram.

Mass spectrometry. All MALDI spectra were obtained on a VT2000 (Vestec Corp., Houston, TX) linear time-of-flight mass spectrometer equipped with a nitrogen laser (337 nm, 3 ns pulse). The accelerating voltage in the ion source was 30 kV. Data were acquired with a transient recorder with 5 ns resolution. Mass resolution ($m/\Delta m$) ranged from 300 to 600 (full width at half-maximum) depending on the sample and laser power. Three candidates for the MALDI matrix, α-cyano-4-hydroxycinnamic acid, sinapinic acid, and 2,5-dihydroxybenzoic acid were evaluated. The matrix selected for this study was α-cyano-4-hydroxycinnamic acid which was found to give the highest sensitivity for our model phosphopeptides, and many other peptides (37), at picomole to subpicomole levels. The matrix was dissolved in aq 33% (v/v) acetonitrile to give a saturated solution at room temperature (~20 mM). To prepare the sample, 1 μl of the solution (0.1% aq TFA:acetonitrile = 1:1, v/v) containing the peptide sample was added to 1 μl of the matrix solution and applied to a flat stainless-steel probe tip. The mixture was then allowed to air dry before being introduced into the mass spectrometer. In certain cases, the analyte/matrix solution was desalted by cation exchange resin (Bio-Rad, AG 50W-X8, mesh size 100–200). Five beads of cation exchange resin were added to a peptide/matrix solution (2 ~ 10 μl) and kept at room temperature for 5 min, then the desalted supernatant solution was applied to the probe tip for mass spectral analysis. Each spectrum was produced by accumulating data from 64 laser shots. Spectra were obtained from different regions of the probe tip, and a representative spectrum was selected for analysis. Time-to-mass conversion was achieved by either external or internal calibration using peaks for Na^+ (m/z 22.99), K^+ (m/z 38.96), matrix peaks (α-cyano-4-hydroxycinnamic acid, $[M + H]^+$ at m/z 190.17 and m/z 379.34 for $[2M + H]^+$), and insulin peaks ($[M + H]^+$ at m/z 5734.6 and m/z 2867.8 for $[M + 2H]^{2+}$).

Mass mapping. A computer program[3] was written using Microsoft Visual Basic Version 1.0 for Windows (Microsoft Corp., Redmond, WA) to calculate the masses of possible peptide fragments and phosphopeptide fragments from the specific enzymatic or chemical degradation of a protein and the m/z value of the mass spectral peak for the corresponding $[M + H]^+$ ion. Because of its limited resolution, MALDI-TOF mass spectrometry provides only the average masses of ions rather than their monoisotopic masses. In the program, the average masses were used instead of monoisotopic masses which are typically used by programs designed for MS analysis (e.g., MacProMass software, by Terry Lee and Sunil Vemuri, Beckman Research Institute of the City of Hope, Duarte, CA). Because partial digestion is not an uncommon situation even when long incubation times are used, the program computes masses for peptides from not only complete but also partial digestion.

[3] This program is available upon request from the authors.

RESULTS

Three model peptides were phosphorylated at serine, threonine, and tyrosine, respectively. KRPpSQRHGSKY-amide, KRpTLRR, and LKRApYLG-amide were synthesized to study the validity of identifying a phosphopeptide by MALDI-TOF mass spectrometry. The first step is to determine whether the peptide is phosphorylated. This can be done by obtaining MALDI mass spectra before and after treatment with alkaline phosphatase. Figure 1 shows the expected −80-Da mass shifts in the MALDI spectra due to the loss of phosphate moieties after dephosphorylation, confirming that these three model peptides were phosphorylated. The dephosphorylation reactions were carried out on the probe by treating 1 pmol of each phosphopeptide with alkaline phosphatase. The MALDI signal was then obtained on the same probe sample. The high concentration of buffer used often made the dried sample form a thick cake. In some experiments, this hindered signal generation. Moving the laser beam to find the proper spot on the sample surface to obtain optimum signal is often necessary. If problems are encountered at this point, one can remove buffer salts by dipping the sample probe in cold water for 1 or 2 s (39). This simple procedure often gives surprising improvement in the ability to obtain the mass spectrum.

Usually, 1 pmol is the least amount of sample that can be easily handled (40). The lower limits of detectability were examined by serial dilution of the three model phosphopeptides and their dephosphorylated forms; the results are shown in Table 1. The limits of detection of dephosphorylated peptides were not investigated at levels lower than 1 fmol. At the low femtomole level, moving the laser beam to find the proper spot on the target to obtain the optimum signal was usually necessary. Figure 2 shows the mass spectrum of a phosphopeptide, KRPpSQRHGSKY-amide, using only 10 fmol of the analyte. A very good signal-to-noise ratio (~25:1) is still obtained. Compared to FAB, MALDI is clearly much more sensitive. FAB usually requires hundreds of pmols or more of peptide to yield satisfactory signals. Among the three model phosphopeptides, KRPpSQRHGSKY-amide gave the weakest response in FAB-MS. Even 1 nmol of this phosphopeptide in glycerol matrix did not give any detectable signal using a JEOL HX-110 double focusing mass spectrometer.

Because there is only one possible phosphorylation site in the known sequence of this peptide, the presence of a −80-Da mass shift in the spectra of KRTLRR (Figs.

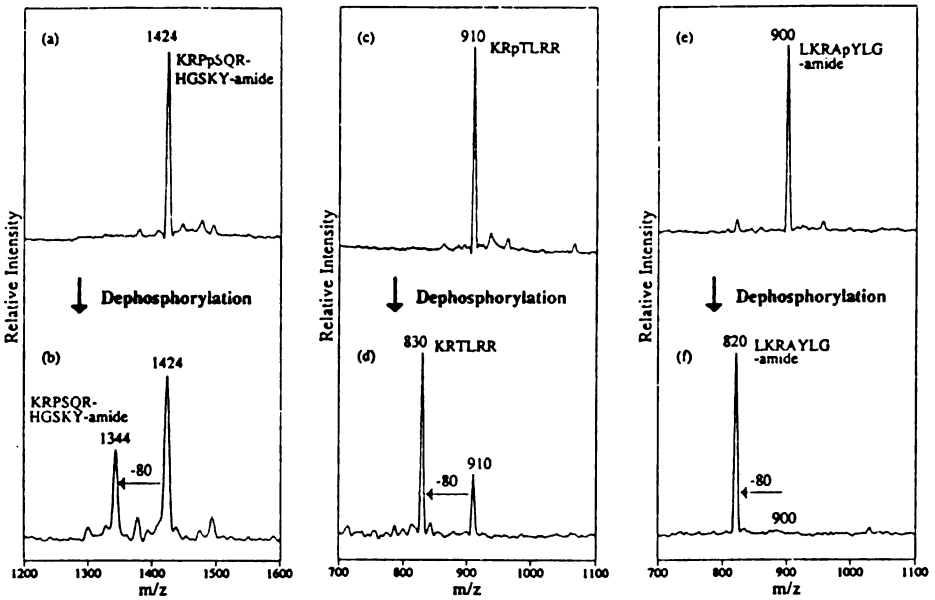


FIG. 1. The −80 Da mass shift of mass spectral peaks due to the loss of phosphate moiety after the dephosphorylation of three model phosphopeptides is shown. The dephosphorylation reaction was done on the probe at room temperature for 6 h by treating phosphopeptide with alkaline phosphatase using 1 pmol of peptide. The MALDI signal was then obtained from the same probe.

1c and 1d) is all the information needed to assign the phosphate to T3. However, in the case of p-KRPSQR-HGSKY-amide, the single −80-Da mass shift upon dephosphorylation (Figs. 1a and 1b) indicates monophosphorylation, but no information on the site of phosphorylation among the three possible sites, S4, S9, and Y11. In such cases, the possible phosphorylation sites can be distinguished by enzymatically or chemically cleaving the peptide into smaller pieces which can then be analyzed by mass spectrometry. Figure 3 shows the tryptic digestion of p-KRPSQRHGSKY-amide which was performed to successfully locate the phosphorylation site among two possible tryptic fragments. Figure 3a shows the peak representing the ionized intact molecule at *m/z* 1424. Following digestion, two peaks appear at *m/z* 591 (which contains S9 and Y11) and *m/z* 852 (which contains S4) (Fig. 3b). Dephosphorylation does not affect the peak at *m/z* 591; however, the peak at *m/z* 852 shifts by −80-Da to *m/z* 772 (Fig. 3c). This information is sufficient to establish S4 phosphorylation.

MALDI is a suitable technique for optimizing enzymatic or chemical reaction conditions because of its high sensitivity and speed of analysis. To monitor the

TABLE 1

The Detectabilities of Three Phosphopeptides and Their Dephosphorylated Forms

Peptide	Phosphorylated form (fmol)	Dephosphorylated form (fmol)
KRPpSQRHGSKY-amide	10	1
KRpTLRR	10	1
LKRApYLG-amide	100	1

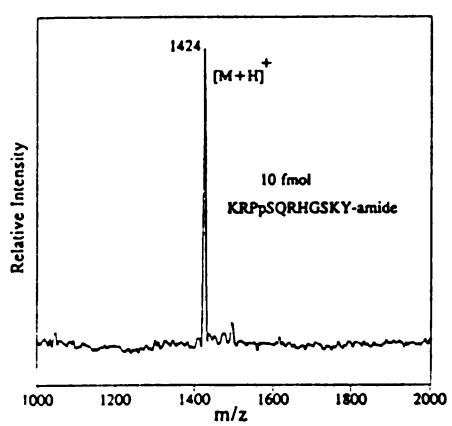


FIG. 2. The molecular ion region of the MALDI spectrum of the phosphopeptide, KRPpSQRHGSKY-amide Only 10 fmol of the analyte was used to obtain this spectrum with good signal-to-noise ratio (~25:1).

extent of reaction, small fractions can be taken from a reaction mixture for direct mass spectrometric analysis. Figure 4 shows the time-course monitoring of the dephosphorylation reaction of KRpTLRR. The results suggest that 15 min of incubation with alkaline phosphatase at 37°C (Fig. 4a) can dephosphorylate this phosphopeptide to the extent that the −80-Da mass shift is easily observed. In Fig. 4a, the peaks which represent both the phosphorylated and dephosphorylated forms of a peptide in the same mass spectrum make the −80-Da mass shift from dephosphorylation easy to detect, even without internal calibration. However, this is not an essential requirement to detect the dephosphorylation reaction when a phosphopeptide is treated with alkaline phosphatase. These results also indicate that the dephosphorylation reaction was complete within 2 h. Another ancillary finding in this study is that a second peptide lacking one arginine residue is present (as indicated by the mass spectral peaks at *m/z* 754 and 674 for the phosphorylated and dephosphorylated forms, respectively). The origin of this component is probably an impurity formed by a missed cycle during the synthesis of KRTLRR,[4] although it could also be due to a contaminating exopeptidase activity in the alkaline phosphatase. This component could not be readily detected by non-mass spectrometric methods.

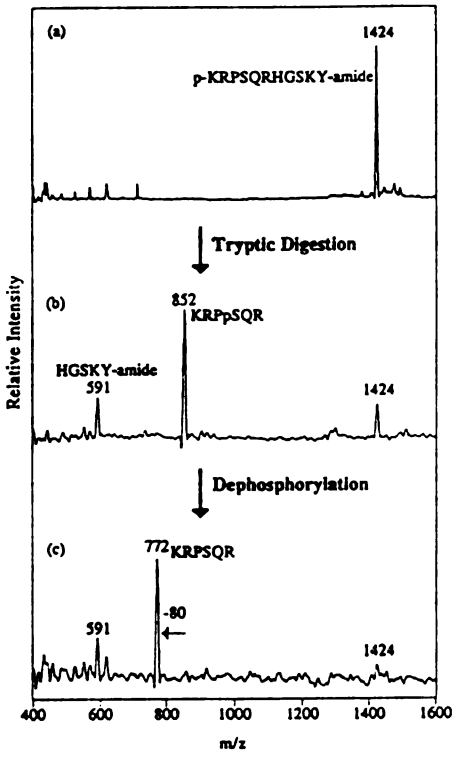


FIG. 3. The tryptic digestion of *p*-KRPSQRHGSKY-amide was performed to locate the phosphorylation site among three possible sites (S4, S9, Y11). 80 μg of phosphopeptide was digested by trypsin. 50 pmol of the digestion mixture was dephosphorylated by alkaline phosphatase and used to obtain each mass spectrum.

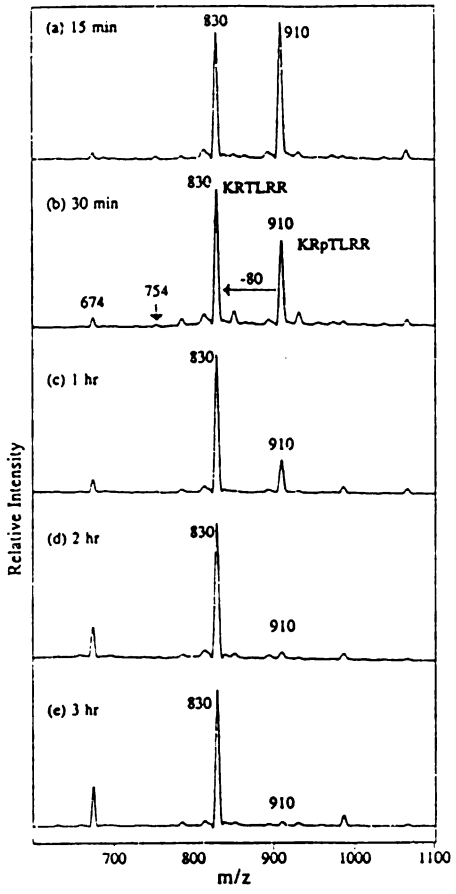


FIG. 4. Time course monitoring of the dephosphorylation reaction of KRpTLRR. 100 pmol of peptide was incubated with alkaline phosphatase at 37°C, 10 pmol of analyte was used to obtain each mass spectrum. The minor peaks at *m/z* 754 and 674 likely correspond to the phosphorylated and dephosphorylated forms, respectively, of a peptide impurity lacking an *R* residue formed during the synthesis of KRpTLRR.

[4] This peptide impurity was also detected in the MALDI spectrum of the original untreated peptide sample, although the impurity peaks (phosphorylated form at *m/z* 754 and unphosphorylated form at *m/z* 674) are below the mass range of the spectra presented in Figs. 1c and 1d.

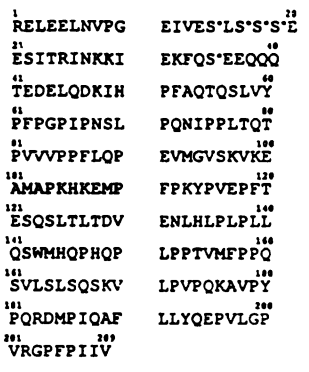


FIG. 5. The primary structure of bovine β-casein.

Bovine β-casein is a well-characterized multiply phosphorylated protein for which the phosphorylation sites have been previously determined by other methods (1,41). Figure 5 shows the primary structure of bovine β-casein (41,42). A 580-μg sample (24 nmol) of bovine β-casein was digested by trypsin (nanomolar amounts were used here, since the focus of this work is not the optimization of digestion reactions, but to demonstrate the utility of MALDI-MS following enzymatic cleavage). Since β-casein is much larger than the previous examples and more products are expected, the analysis was performed after a rough chromatographic separation by reverse-phase HPLC (using 10 μg or 400 pmol digest mixture). Without the aid of radioactive isotope labeling, each HPLC fraction (2 pmol, assuming 100% recovery from HPLC) was subjected to MALDI analysis before and after phosphatase treatment to determine if there were any phosphopeptides in the fraction. The presence of phosphopeptide(s) was reflected by the −80 (or multiples of −80)-Da mass shifts of peaks in the mass spectra following dephosphorylation. Fifteen HPLC fractions were collected; only two phosphopeptides (in fractions 4 and 7) were identified among the tryptic peptides. Figure 6 shows the mass spectra of fractions 4 and 7 which indicated the phosphorylation sites of bovine β-casein. Table 2 lists the predicted masses of peptide fragments between the mass range 0 ~ 4000 Da. All of the major peaks shown in Fig. 6 can be assigned unambiguously to tryptic peptides or phosphopeptides from the predicted masses. From Figs. 6a and 6b and Table 2, a good match between observed and calculated values at *m/z* 2064.0 (calculated 2063.0, before dephosphorylation) and at *m/z* 1983.4 (calculated 1983.0, after dephosphorylation) indicates that only monophosphorylated peptide p33-48, p-FQSEEQQQTEDELQDK, fits the observed data. It was previously determined that phosphorylation does not occur at threonine on bovine β-casein (43). Usually, this information can be obtained by phosphoamino acid analysis (44). Based on this assumption, the monophosphorylated peptide can be uniquely assigned as p33-48, FQpSEEQQQTEDELQDK. Alternatively, enzymatic digestion by another specific protease can be performed, so that the mass mapping information from two or more different specific degradations is combined to indicate the site of phosphorylation. HPLC fraction 4, which contains phosphopeptide p33-48, FQpSEEQQQTEDELQDK, also contains another peptide, 106-113, HKEMPFPK (*m/z* 1014.1). This shows that the method used here can tolerate coeluting components in an HPLC fraction, in contrast to the poor tolerance for peptide impurities in Edman degradation analysis. This advantage greatly lowers the criteria for the chromatographic separation of the peptides.

For fraction 7, the −320-Da mass shift (from *m/z* 3125.2 to *m/z* 2804.8) after alkaline phosphatase treatment shown in Figs. 6c and 6d suggests that there are four phosphorylation sites on this peptide fragment. By referring to Table 2, this phosphopeptide is uniquely assigned to be pppp1-25, pppp-RELEELNVPGEIVESLSSSEESITR. The peak at *m/z* 3125.2 in Fig. 6c could correspond to either a peptide of calculated mass 3124.0 (differing by 0.04%) or 3127.8 (differing by 0.08%) in Table 2. However, the latter requires six phosphorylated residues and can thus be excluded because alkaline phosphatase treatment indicated that this peptide has only four phosphorylated residues. That is, if this peptide does have six phosphorylation sites, the partially dephosphorylated peak at *m/z* 2804.8 (=3125.2 − 80 × 4) would be unlikely. The additional peak at *m/z* 2648.8 (Fig. 6d) probably represents the dephosphorylated form of tryptic fragment 2-25. Presumably incomplete separation gave a mixture of 1-25 and 2-25, which were both dephosphorylated. The peptide 1-25, RELEELNVPGEIVESLSSSEESITR, has five serine residues and one threonine residue, four of which are phosphorylated. To further characterize the phosphorylation sites on this peptide, the enzymatic digestion of HPLC fraction 7 by endoproteinase Glu-C was performed. A predicted peak at *m/z* 1059.8 was found in the mass spectrum (Fig. 6e), which corresponded to the sequence, pppp15-21, pSLpSpSpSEE. This peak provides unambiguous information about the remaining phosphorylation sites. Only five peaks were predicted to be found in the *m/z* 1000 ~ 1100 range: *m/z* 1000.1 (peptide fragment 3-11), *m/z* 1031.0 (peptide fragment p12-20), *m/z* 1058.7 (peptide fragment pppp15-21), *m/z* 1080.1 (peptide fragment 12-21), and *m/z* 1099.2 (peptide fragment 5-14), each of which could be easily distinguished without ambiguity. With this information, the phosphopeptide observed in Fig. 6c with a peak at *m/z* 3125.2 was assigned as pppp1-25, RELEELNVPGEIVEpSLpSpSpSEESITR. These conclusions concerning phosphory-

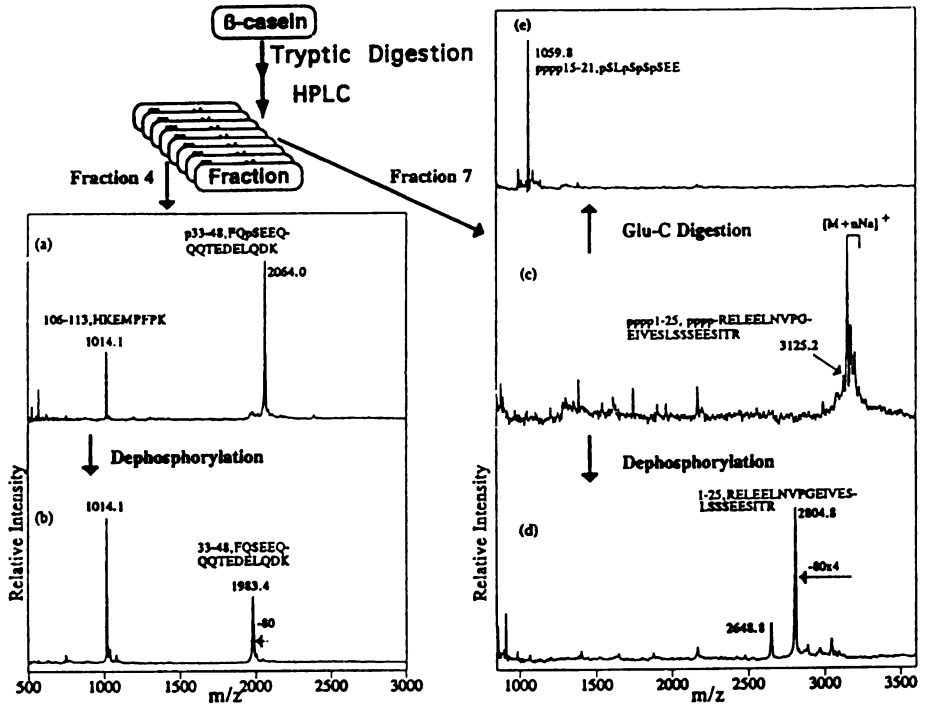


FIG. 6. The mass spectra of HPLC fractions 4 and 7 which indicated the phosphorylation sites of bovine β-casein. 2 pmol (assuming 100% recovery from HPLC) of each fraction was incubated with alkaline phosphatase at 37°C for 15 min and used to obtain each mass spectrum. 100 pmol (assume 100% recovery from HPLC) of HPLC fraction 7 was digested with endoproteinase Glu-C at room temperature for 4 h, and 10 pmol of analyte was used to obtain the mass spectrum.

lation sites of bovine β-casein coincide with the five phosphorylated sites reported previously (41). Table 3 summarizes our data and the assignments of the peaks in the spectra.

From Fig. 6c, it can be seen that the peptide pppp1-25, RELEELNVPGEIVEpSLpSpSpSEESITR, has a strong tendency to form multiple-sodium adducts in MALDI. Cation exchange resin (AG 50W-X8, Bio-Rad) was used to desalt the analyte/matrix solution before loading sample to probe tip for mass spectral analysis (45). Figure 7 shows the mass spectrum of HPLC fraction 7 of which analyte/matrix solution has been desalted by cation exchange resin. Most of the sodium adducts of this peptide were eliminated by this method. Only the peak for $[M + Na]^+$ still remained intense, although the peak for $[M + H]^+$ dominates. This also qualitatively indicates that this multiply phosphorylated peptide exhibits a very strong affinity for Na^+.

It would be highly desirable to analyze the digestion mixture by MALDI without any chromatographic separations. This direct approach was evaluated. Figure 8 shows the mass spectrum of tryptic digestion mixture of bovine β-casein before and after alkaline phosphatase treatment. Table 4 lists the identified peaks observed in the MALDI spectra. Most of the peaks shown in the spectrum can be interpreted as tryptic peptides. However, there are still some unidentifiable peaks which might be caused by impurities in the sample or nonspecific cleavages. From the presence of tryptic peptides 98-107, 98-113, 114-169, and 106-169 shown in Table 4, it is clear that the enzymatic cleavage reaction did not go to completion. However, it is likely that peptides 1-25, 29-32, 33-48, 49-97, and 177-183 are present in comparable amounts. If all tryptic peptides give the same response in MALDI, one should expect that the signals from these peptides would have comparable intensities in the spectrum. Nevertheless, the signal intensities vary considerably for these tryptic peptides as shown in Fig. 8. The peaks at m/z 2911.7 and 2804.2 are dominant in Figs. 8a and 8b, respectively. Presumably, such pronounced differences in response can be attributed to different ionization efficiencies of peptides by MALDI and/or suppression effects. We have previously reported that the proton affinities of peptides might af-

TABLE 2

Possible Tryptic Peptides (Includes All Possible Phosphopeptides by Considering Phosphorylation at Serine, Threonine, and Tyrosine Residues) from Bovine β-Casein in the Mass Range of 0 ~ 3500 Da Calculated from the Mass-Mapping Program

m/z	Peptide	*m/z*	Peptide
147.2	29-29	2887.8	3p, 2-25
175.2	1-1	2911.5	184-209
246.3	98-99	2917.0	1p, 26-48
284.3	106-107	2964.0	2p, 1-25
374.4	26-28	2967.8	4p, 2-25
389.4	30-32	2991.5	1p, 184-209
502.6	26-29	2997.0	2p, 26-48
517.6	29-32	2999.5	177-202
646.8	100-105	3003.3	2-28
742.9	203-209	3044.0	3p, 1-25
748.9	108-113	3047.8	5p, 2-25
781.0	170-176	3079.5	1p, 177-202
830.9	177-183	3083.3	1p, 2-28
873.1	26-32	3124.0	4p, 1-25
874.1	98-105	3127.8	6p, 2-25
910.9	1p, 177-183	3131.4	2-29
912.1	100-107	3159.5	1-28
1014.2	106-113	3159.5	2p, 177-202
1139.4	98-107	3163.3	2p, 2-28
1592.9	170-183	3204.0	5p, 1-25
1642.0	100-113	3211.4	1p, 2-29
1672.9	1p, 170-183	3239.5	1p, 1-28
1869.3	98-113	3243.3	3p, 2-28
1983.0	33-48	3284.0	6p, 1-25
2063.0	1p, 33-48	3287.6	1-29
2143.0	2p, 33-48	3291.4	2p, 2-29
2187.6	184-202	3319.5	2p, 1-28
2267.6	1p, 184-202	3323.3	4p, 2-28
2353.4	30-48	3367.6	1p, 1-29
2433.4	1p, 30-48	3371.4	3p, 2-29
2481.6	29-48	3399.5	3p, 1-28
2513.4	2p, 30-48	3403.3	5p, 2-28
2561.6	1p, 29-48	3447.6	2p, 1-29
2641.6	2p, 29-48	3451.4	4p, 2-29
2647.8	2-25	3479.5	4p, 1-28
2727.8	1p, 2-25	3483.3	6p, 2-28
2804.0	1-25		
2807.8	2p, 2-25		
2837.0	26-48		
2884.0	1p, 1-25		

fect proton transfer chemistry which occurs in the MALDI sources (46). Another important observation is that the peak at *m/z* 2911.7 representing the nonphosphorylated tryptic fragment 184-209 is very intense (base peak), but its intensity was reduced after the digestion mixture was treated with alkaline phosphatase. In the phosphatase-treated sample spectrum (Fig. 8b), a new peak at *m/z* 2804.2, originating from the dephosphorylated tryptic peptide, pppp1-25, RELEELNVPGEIVESLSSSEESITR, emerged as the base peak, although the tryptic fragment 184-209 was not modified by the enzymatic reaction (alkaline phosphatase). The ionization of the peptide corresponding to the peak at *m/z* 2911.7 may be suppressed by the peptide contributing the peak at *m/z* 2804.2 or these peptides may have dramatically different responses in MALDI analysis. Thus, for mixtures, all peptide fragments are not necessarily detectable in MALDI-MS (37,47,48).

DISCUSSION

There are three steps to locate phosphorylation sites in a phosphoprotein by the proposed method: (i) degrade a phosphoprotein into small peptides and phosphopeptides by specific enzymatic or chemical reactions; (ii) identify phosphopeptides by −80 (or multiples of −80)-Da mass shifts after dephosphorylation with alkaline phosphatase; (iii) locate phosphorylation sites by mass mapping. The latter step may involve the use of several specific proteases and subsequent mass spectral analyses by MALDI-MS. Like other methods to locate phosphorylation sites, phosphoproteins must first be cleaved into small peptides and phosphopeptides by enzymatic or chemical reactions prior to analysis. The digestion mixture of peptides and phosphopeptides from a phosphoprotein can be fractionated by reverse-phase HPLC, and then subjected to MALDI analysis. Direct analysis of the unfractionated digestion mixture by MALDI is also possible. However, there are still practical difficulties encountered at times when a complicated digestion mixture is analyzed due to the differential responses and/or suppression of peptides in a mixture. A reagent is then required which removes the phosphate moiety from a phosphopeptide, so a phosphopeptide can be identified from the −80-Da mass shift ($-OPO_3H_2 \rightarrow -OH$). Enzymatic dephosphorylation by a phosphatase is useful for this purpose. If there is more than one phosphorylation site in the peptide, then mass shifts of multiples of −80 Da will be observed. With a known protein sequence, locating phosphorylation sites is achieved by "mass mapping." Owing to the advances in techniques of molecular biology, solving the primary structure of a protein has become much easier. To fulfill the conditions of mass mapping, one requirement must be met: the phosphoprotein must be cleaved in a specific way, so the number of fragments and their corresponding masses can be predicted and calculated. Once a list of possible fragments and their corresponding masses is made, a peptide fragment can be readily mapped by its mass to find its identity. In cases when more than one peptide may be close in mass, secondary enzymatic digests and MALDI analysis can be used. Various endoproteinases, such as trypsin, endoproteinase Glu-C, and endoproteinase Asp-N, which cleave proteins at specific sites, are suitable for this purpose. Specific degradation by chemical reagents, such as cyanogen bromide, would also be useful, as long as the masses of generated peptide fragments are predictable.

TABLE 3

The Assignment of Major Peaks Shown in Figs. 6a-6e

Peptide (M)	Figure	Observed *m/z*	Calculated *m/z* for $[M + H]^+$
p33-48, FQpSEEQQQTEDELQDK	6a	2064.0	2063.0
33-48, FQSEEQQQTEDELQDK	6b	1983.4	1983.0
106-113, HKEMPFPK	6a,b	1014.1	1014.2
ppppl-25, pppp-RELEELNVPGEIVESLSSSEESITR	6c	3125.2	3124.0
1-25, RELEELNVPGEIVESLSSSEESITR	6d	2804.8	2804.0
ppppl5-21, pSLpSpSpSEE	6e	1059.8	1058.7

The major advantages of the proposed method are: fast analysis, easy operation, high sensitivity, and ability to analyze mixtures. Importantly, there is no necessity for radioactive isotope labeling, a requirement for other procedures (2–18). In our laboratory, 24 samples can be readily analyzed within 2 h. Once the samples are prepared on probes, each sample requires only 1 min to acquire a spectrum. The high sensitivity of this method is also important when the amount of available phosphoprotein is limited. The low to subpicomole-level sensitivity allows, in many cases, the use of multiple enzymatic digests to obtain the desired information. The ability to analyze mixtures eliminates tedious separation work especially given the likelihood that two or more components will coelute under typical HPLC conditions.

Another question associated with this method is: what is the mass accuracy requirement for the mass spectrometer in this approach? The mass accuracy of MALDI-TOF mass spectrometer has been reported to be as good as 0.01% by several research groups (31,32). Such mass accuracy corresponds to 0.1 mass unit for an ion of mass 1000 Da, or 1 mass unit for an ion of mass 10,000 Da. However, in our experience, a mass accuracy of 0.1% is a value which can be realistically maintained for routine analyses. To determine if a mass accuracy of 0.1% is sufficient for the mass mapping of the phosphorylation sites in a protein, we have performed some calculations using the computer program described in

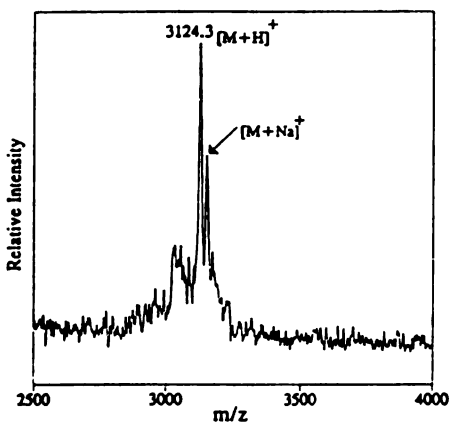


FIG. 7. The mass spectrum of HPLC fraction 7, for which the analyte/matrix solution was desalted using a cation exchange resin. 20 pmol fraction 7 was mixed with matrix solution and five beads of cation exchange resin at room temperature for 5 min, then the desalted solution was applied to the probe tip for mass spectral analysis.

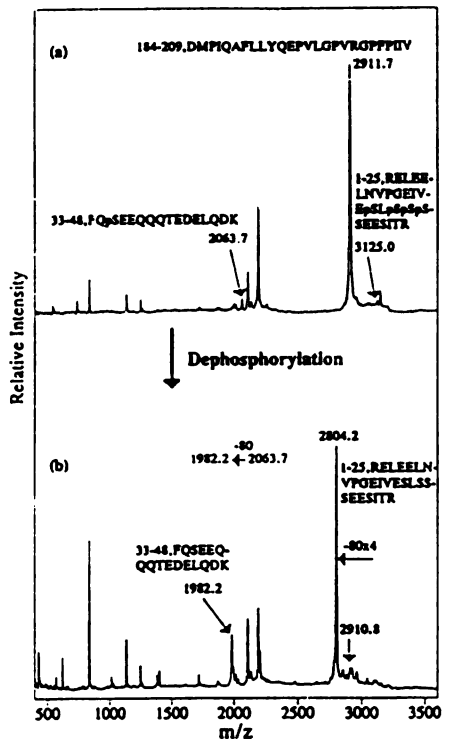


FIG. 8. The mass spectra of the tryptic digestion mixture of bovine β-casein before and after the phosphatase treatment. 20 pmol of the digestion mixture was incubated with alkaline phosphatase at 37°C for 15 min. 2 pmol of the analyte was used to obtain each mass spectrum.

TABLE 4

Identified Peaks of Tryptic Peptides Observed in Figs. 8a and 8b

Peptide (M)	Observed m/z	Calculated m/z for $[M + H]^+$
29-32	538.8[a]	539.6[a]
177-183	831.7	830.9
98-107	1138.5	1139.4
98-113	1868.6	1869.3
33-48	1982.2[b]	1983.0
1p, 33-48	2063.7	2063.0
184-202	2187.9	2187.6
1-25	2804.2[b]	2804.0
184-209	2911.7	2911.5
4p, 1-25	3125.0	3124.0
	3149.5[a]	3146.0[a]
49-97	5314.1[c]	5320.2
114-169	6361.7[c]	6363.3
106-169	7359.0[c]	7358.5

[a] Sodium adduct.
[b] Data from Fig. 8b. The rest of data were obtained from Fig. 8a.
[c] Peak not shown in Fig. 8.

this paper. The computer program can generate a list of possible peptide fragments from a specific enzymatic digestion of a protein and sort them by their mass values. The difference (Δ_i) and % relative difference ($\%\Delta_i$) of mass values between two adjacent peptides in the list are defined as,

$$\Delta_i = (\text{mass}_{i+1}) - (\text{mass}_i), \text{ and}$$
$$\%\Delta_i = 100\% \times \Delta_i/(\text{mass}_i),$$

where, i = 1 to $(n - 1)$, and n is the number of peptide fragments in the list. For example, in Table 2, the pair of ions m/z 2143.0 and 2063.0 yields a $\%\Delta_i$ value of 3.9%; while the pair of ions m/z 2999.5 and 2997.0 yields a $\%\Delta_i$ value of 0.08%. The $\%\Delta_i$ of tryptic peptide fragments from bovine β-casein were calculated and categorized into two subsets:

I. $\%\Delta_i$ is less than 0.1%.

II. $\%\Delta_i$ is greater than or equal to 0.1%.

Since the mass accuracy of MALDI-TOF is often better than 0.1%, the mass assignment of a peptide$_i$ in mass mapping would be unambiguous if both $\%\Delta_{i-1}$ and $\%\Delta_i$ fall into subset II (that is, the peak for peptide$_i$ could be separated from the closest peaks above and below on the m/z axis). For the tryptic peptides from bovine β-casein (MW 23,983), the numbers of peptides with $\%\Delta_i$ values greater than (or equal to) 0.1% and less than 0.1% were calculated and plotted as a function of mass (Fig. 9). Two situations were considered: all peptides with possible phosphorylation at serine, threonine, and/or tyrosine (Fig. 9a) and all possible peptide fragments after complete dephosphorylation (Fig. 9b).

From Fig. 9a, it is clear that to identify a peak representing a potential phosphopeptide which has not been treated with alkaline phosphatase in the mass range 0 ~ 2000 Da by mass mapping is unambiguous (24 of 24 peptides with $\%\Delta_i$ greater than or equal to 0.1%). For a peak in the mass range higher than 2000 Da, it is more ambiguous. For example, the peak at m/z 2064.0 in Fig. 6a can be unambiguously mapped as tryptic peptide p33-48, p-FQSEEQQQTEDELQDK by using Table 2. In contrast, the peak at m/z 3125.2 in Fig. 6c cannot be unambiguously mapped. The calculated masses of two tryptic peptides (m/z 3124.0, ppppl-25, and m/z 3127.8, pppppp2-25) are both close to m/z 3125.2 and differ by less than 0.1% in mass. However, the ambiguity can be clarified by treating this peptide with alkaline phosphatase to reveal the number of phosphate groups on it. When one has identified a phosphopeptide by a mass shift due to dephosphorylation, it is only necessary to match the dephosphorylated form of this peptide among all tryptic peptides without considering phosphorylation. From Fig. 9b, it is found that there is no ambiguity in mass mapping in the mass range 0 ~ 8000 Da, and the chance of finding fragments with $\%\Delta_i$ less than 0.1% is only 3 out of 126 peptides in the mass range 0 ~ 24000 Da. However, because the primary object is to locate the phosphorylation sites within a small domain

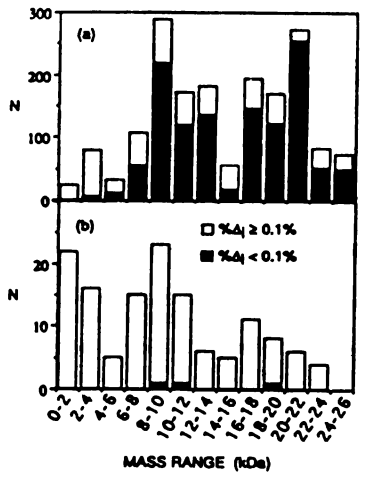


FIG. 9. The number of possible tryptic peptides (N) from bovine β-casein were calculated for different mass ranges. The proportion of the total possible tryptic peptides from bovine β-casein with $\%\Delta_i$ less than 0.1% (black) and $\%\Delta_i$ greater than or equal to 0.1% (white) are indicated for (a) the case which there is possible phosphorylation at serine, threonine, and/or tyrosine, and (b) the case which all peptides have been dephosphorylated. See text for discussion.

along the polypeptide chain, a peptide fragment smaller than mass 3000 Da, or 5000 Da at most, will yield useful information. It is concluded that the 0.1% mass accuracy is sufficient for mass mapping to identify the tryptic peptides from β-casein, and thus locate its phosphorylation sites.

A major limitation of MALDI-TOF mass spectrometry is that the signal intensities vary considerably between peptides, particularly when they are present in a mixture. Some peptides may not give a detectable signal (46). This can hamper the proposed method in certain cases. In addition, suppression effects have been observed in a complicated tryptic digest mixture, which potentially make interpretation of MALDI spectra difficult. Thus, sample fractionation may still be required in some cases. Sample consumption in this method is primarily limited by sample handling in digestion and separation, rather than by the detection limits of MALDI-TOF-MS. The incorporation of microdigestion and microseparation into this method is currently under investigation in our laboratory.

ACKNOWLEDGMENT

This work was supported by the Biomedical Resource Technology Program of the National Center for Research Resources of National Institutes of Health (RR-00480 to J. T. Watson).

REFERENCES

1. Roach, P. J (1991) *J. Biol. Chem.* **266**, 14139-14142.
2. Cleveland, D. W., Fischer, S. G., Kirschner, M. W., and Laemmli, U. K. (1977) *J. Biol. Chem* **252**, 1102-1106.
3. Juhl, H., and Soderling, T. R. (1983) *in* Methods in Enzymology (Corbin, J. D., and Hardman, J. G., Eds.), Vol. 99, pp. 37-48, Academic Press, San Diego.
4. Boyle, W. J., Geer, P. V. D., and Hunter, T. (1991) *in* Methods in Enzymology (Hunter, T., and Sefton, B. M., Eds.), Vol. 201, pp. 110-149, Academic Press, San Diego.
5. Campbell, D. G., Hardie, D. G., and Vulliet, P. R. *J. Biol. Chem.* **261**, 10489-10492.
6. Meyer, H. E., Hoffmann-Posorske, E., Korte, H., and Heilmeyer, L. M. G., Jr., (1986) *FEBS Lett.* **204**, 61-66.
7. Aebersold, R., Watts, J. D., Morrison, H. D., and Bures, E. J. (1991) *Anal. Biochem.* **199**, 51-60.
8. Holmes, C. F. B. (1987) *FEBS Lett.* **215**, 21-24.
9. Meyer, H. E., Hoffmann-Posorske, and Heilmeyer, L. M. G., Jr., (1991) *in* Methods in Enzymology (Hunter, T., and Sefton, B. M., Eds.), Vol. 201, pp. 169-185, Academic Press, San Diego.
10. Haystead, T. A. J., Campbell, D. G., and Hardie, D. G. (1988) *Eur. J. Biochem* **175**, 347-354.
11. Poulter, L., Ang, S.-K., Gibson, B. W., Williams, D. H., Holmes, C. F. B., Caudwell, F. B., Pitcher, J., and Cohen, P. (1988) *Eur. J. Biochem.* **175**, 497-510.
12. Chirat, F., Martinage, A., Briand, G., Kouach, M., Van Dorsselaer, A., Loir, M., and Sautiere, P. (1991) *Eur. J. Biochem.* **198**, 13-20.
13. May, L. T., and Sehgal, P. B. (1992) *Biochem. Biophys. Res Commun.* **185**, 524-530.
14. Holmes, C. F. B., Tonks, N. K., Major, H., and Cohen, P. (1987) *Biochim. Biophys. Acta* **929**, 208-219.
15. Xu, Z.-S., Liu, W.-S., and Willard, M. B. *J. Biol. Chem.* **267**, 4467-4471.
16. Carrey, E. A., Campbell, D. G., and Hardie, D. G. (1985) *EMBO J.* **4**, 3735-3742.
17. Wettenhall, R. E. H., Aebersold, R. H., and Hood, L. E. (1991) *in* Methods in Enzymology (Hunter, T., and Sefton, B. M., Eds.), Vol. 201, pp. 186-199, Academic Press, San Diego.
18. Bodwell, J. E., Orti, E., Coull, J. M., Pappin, D. J. C., Smith, L. I., and Swift, F. (1991) *J. Biol. Chem.* **266**, 7549-7555.
19. Labdon, J. E., Nieves, E., and Schubart, U. K. (1992) *J. Biol. Chem.* **267**, 3506-3513.
20. Luca, M. D., Graan, P. N. E., Angelis, L. D., Gispen, W. H., and Cattabeni, F. (1992) *FEBS Lett.* **301**, 150-154.
21. Amess, B., Manjarrez-Hernandez, H. A., Howell, S. A., Learmonth, M., and Aitken, A. (1992) *FEBS Lett.* **297**, 285-291.
22. Gibson, B. W., and Cohen, P. (1990) *in* Methods in Enzymology (McCloskey, J. A., Ed.), Vol. 193, pp. 480-501, Academic Press, San Diego.
23. Poulter, L., Ang, S.-K., Williams, D. H., and Cohen, P. (1987) *Biochim. Biophys. Acta* **929**, 296-301.
24. Gibson, B. W., Falick A. M., Burlingame, A. L., Nadasdi, L., Nguyen, A. C., and Kenyon, G. L. (1987) *J. Am. Chem. Soc.* **109**, 5343-5348.
25. Petrilli, P., Pucci, P., Morris, H. R., and Addeo, F. (1986) *Biochem. Biophys. Res. Commun.* **140**, 28-37.
26. Fenselau, C., Heller, D. N., Miller, M. S., and White, H. B., III, (1985) *Anal. Biochem.* **150**, 309-314.
27. Cohen, P., Gibson, B. W., and Holmes, C. F. B. (1991) *in* Methods in Enzymology (Hunter, T., and Sefton, B. M., Eds.), Vol 201, pp. 153-169, Academic Press, San Diego.
28. Erickson, A. K., Payne, D. M., Martino, P. A., Rossomando, A J., Shabanowitz, J., Weber, M. J., Hunt, D. F., and Sturgill, T. W. (1990) *J. Biol. Chem.* **265**, 19728-19735.
29. Michel, H., Hunt, D. F., Shabanowitz, J., and Bennett, J. (1988) *J. Biol. Chem.* **263**, 1123-1130.
30. Karas, M., Bachmann, D., Bahr, U., and Hillenkamp, F. (1987) *Int. J. Mass Spectrom. Ion Processes* **78**, 53-68.
31. Karas, M., Bahr, U., and Giebmann, U. (1991) *Mass Spectrom. Rev.* **10**, 335-357.
32. Hillenkamp, F., Karas, M., Beavis, R. C., and Chait, B. T. (1991) *Anal. Chem.* **63**, 1193A-1202A.
33. Nuwaysir, L. M., and Stults, J. T. (1993) *J. Am. Soc. Mass Spectrom.* **4**, 662-669.
34. Huddleston, M. J., Annan, R. S., Bean, M. F., and Carr, S. A. (1993) *J. Am. Soc. Mass Spectrom.* **4**, 710-717.
35. Ding, J., Burkhart, W., and Kassel, D. B. (1994) *Rapid Commun. Mass Spectrom.* **8**, 94-98.
36. Yip, T.-T., and Hutchens, W. (1992) *FEBS Lett.* **308**, 149-153.
37. Wang, K., Liao, P.-C., Allison, J., Gage, D., Andrews, P. C., Lubman, D., Hanash, S. M., and Strahler, J. R. (1993) *J. Biol. Chem.* **268**, 14269-14277.
38. Andrews, D. M., Kitchen, J., and Seale, P. N. (1991) *Int. J. Peptide Protein Res.* **38**, 469-475.

39. Beavis, R. C., and Chait, B. T. (1990) *Anal. Chem.* **62**, 1836–1840.
40. Suelter, C. H., and Deluca, M. (1983) *Anal. Biochem.* **135**, 112–119.
41. Dumas, B. R., Brignon, G., Grosclaude, F., and Mercier, J.-C. (1972) *Eur. J. Biochem.* **25**, 505–514.
42. Stewart, A. F., Bonsing, J., Beattie, C. W., Shah, F., Willis, I. M., and Mackinlay, A. G. (1987) *Mol. Biol. Evol.* **4**, 231–241.
43. Dumas, B. R., Brignon, G., Grosclaude, F., and Mercier, J.-C. (1971) *Eur. J. Biochem.* **20**, 264–268.
44. Hunter, T., and Sefton, B. M., Eds. (1991) Methods in Enzymology, Vol. 201, pp. 3–101, Academic Press, San Diego.
45. Nordhoff, E., Ingendoh, A., Cramer, R., Overberg, A., Stahl, B., Karas, M., Hillenkamp, F., and Crain, P. F. (1992) *Rapid Commun. Mass Spectrom.* **6**, 771–776.
46. Liao, P.-C., and Allison, J., Investigating the Ionization Step in MALDI: Matrix-Dependent Formation of $[M + H]^+$ *vs* $[M + Na]^+$ Ions of Small Peptides; Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics, San Francisco, CA, May 30–June 4, 1993.
47. Billeci, T. M., and Stults, J. T. (1993) *Anal. Chem.* **65**, 1709–1716.
48. Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., and Watanabe, C. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 5011–5015.

Appendix V

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 268, No. 19, Issue of July 5, pp. 14269-14277, 1993
Printed in U.S.A.

Phorbol 12-Myristate 13-Acetate-induced Phosphorylation of Op18 in Jurkat T Cells

IDENTIFICATION OF PHOSPHORYLATION SITES BY MATRIX-ASSISTED LASER DESORPTION IONIZATION MASS SPECTROMETRY*

(Received for publication, November 30, 1992, and in revised form, February 26, 1993)

Y. Karen Wang‡§¶, Pao-Chi Liao‖, John Allison, Douglas A. Gage‖, Philip C. Andrews‡‡, David M. Lubman§, Samir M. Hanash‡, and John R. Strahler‡§§**

*From the Departments of ‡Pediatrics, ‡‡Biochemistry, and §Chemistry, University of Michigan, School of Medicine, Ann Arbor, Michigan 48109-0510, and Departments of ‖Biochemistry and **Chemistry, Michigan State University, East Lansing, Michigan 48824*

Op18 is a widely expressed, cell cycle-regulated, phosphoprotein involved in signal transduction of a variety of stimuli. In actively proliferating Jurkat T cells which express Op18 at high level, phorbol 12-myristate 13-acetate (PMA) treatment induces a rapid increase in the level of several Op18 phosphorylated forms. To determine phosphorylation sites involved in the PMA effect, the major Op18 phosphorylated forms were resolved in Jurkat T cells, before and after treatment with PMA, using preparative immobilized pH gradient-based two-dimensional polyacrylamide gel electrophoresis. Tryptic fragments of phosphorylated Op18 were analyzed by two-dimensional thin layer peptide mapping and were resolved by reverse-phase high performance liquid chromatography prior to analysis by matrix-assisted laser desorption ionization mass spectrometry. Phosphorylation sites were identified by further treatment of the proteolytic fragments with different enzymes and determination of the mass shifts by matrix-assisted laser desorption ionization mass spectrometry. Two major phosphorylation sites were identified. Low constitutive levels of phosphorylation at Ser^{25} and Ser^{38} in Op18a and Op18b was demonstrated. Treatment with PMA resulted in enhanced phosphorylation of Ser^{25} in Op18a and of both Ser^{25} and Ser^{38} in Op18b. Taken together with prior studies of Op18 phosphorylation, the data suggest that Op18 phosphorylation occurs at identical sites in different tissues and organisms.

Phosphorylation of the highly conserved cytosolic protein Op18 has been implicated in signal transduction in a wide variety of cell types (1–6). While the effect of Op18 phosphorylation on activity of the protein is not known, several findings have emerged that suggest an important role for Op18 phosphorylation. Freshly isolated PBL[1] constitutively express low levels of mostly unphosphorylated Op18 (1). Upon activation of resting PBL through the T cell receptor, there is rapid phosphorylation of Op18. At least five phosphorylated forms have been identified that are differentially expressed during the cell cycle (7). In proliferating cells, such as Jurkat leukemic T cells, there is an increase in the level of phosphorylated forms of Op18 in response to treatment with phorbol 12-myristate 13-acetate (PMA). While Op18 phosphorylation following activation of resting T cells is associated with progression through the cell cycle, in Jurkat T cells, phosphorylation of Op18 in response to PMA is associated with down-regulation of cell proliferation (8, 9). Evidence also exists for altered Op18 phosphorylation in pathologic states. In neuroblastoma tumors, the phosphorylation of Op18 is substantially diminished in aggressive tumors in which the N-*myc* oncogene is amplified relative to more benign tumors with a nonamplified N-*myc* oncogene (10). In contrast to proliferating normal lymphoid cells in which Op18 is substantially phosphorylated, in freshly isolated acute leukemia cells, Op18 occurs at high levels, predominantly in a nonphosphorylated form (11).

An understanding of the potential role of Op18 in signal transduction and of the significance of its variable phosphorylation requires elucidation of the mechanism of its phosphorylation in response to various stimuli and of the basis for its altered phosphorylation in pathologic states. Of basic importance is the identification of residues that are phosphorylated and that account for the different phosphorylated forms of Op18. Previous studies have established, based on phosphoamino acid analysis, that phosphorylation occurs on serine residues (6, 12). Several approaches have become available to characterize protein post-translational modifications by mass spectrometry. We have recently demonstrated that Op18 undergoes N-terminal acetylation, on the basis of data obtained by FAB/CAD/MS/MS (13). Studies by others using enzymatic digestion and on-line HPLC FAB/MS have led to

* This work was supported by Department of Energy Grant 87 ER 60533 (to J. R. S.), National Institutes of Health Grant CA32146 (to S. M. H.), with equipment made available in the Michigan State University Mass Spectrometry Facility supported in part by Grant RR0480-23 (To D. A. G., J. A., and P.-C. L.) from the Biomedical Resource Technology Program of the National Center for Research Resources of National Institutes of Health, and National Science Foundation Grant CHE9022610 (to D. M. L.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ This work was performed in partial fulfillment of the requirements for the Ph.D. degree. Current address: Sandoz Pharmaceuticals, 59 Rt. 10, Bldg. 405, E. Hanover, NJ 07936.

§§ To whom correspondence should be addressed: University of Michigan, R4451 Kresge I, Box 0510, Ann Arbor, MI 48109-0510. Tel.: 313-763-9311; Fax: 313-763-2025.

[1] The abbreviations used are: PBL, peripheral blood lymphocytes; PMA, phorbol 12-myristate 13-acetate; FAB, fast atom bombardment; CAD, collisionally activated dissociation; MS, mass spectrometry; HPLC, high performance liquid chromatography; MALDI, matrix-assisted laser desorption ionization; PAGE, polyacrylamide gel electrophoresis; IPG, immobilized pH gradient; PVDF, poly(vinylidene fluoride).

the identification of Ser^{25} and Ser^{36} as phosphorylation sites of Op18 (p19) isolated from bovine brain (14).

MALDI/MS has been used primarily for molecular weight determinations of intact proteins (15–17). The high sensitivity (picomole to sub-picomole detection limits) of this technique and its property of tolerating relatively high concentrations of contaminants such as buffer salts make it well suited for molecular weight determinations of peptides derived from small quantities of cellular proteins. Recently, several groups have used MALDI/MS in combination with enzymatic or chemical modification of peptides to obtain sequence information by analyzing mass shifts in the MALDI spectrum following peptide modification (18–21).

In this investigation we have undertaken an analysis of Op18 phosphorylation in response to treatment of Jurkat T cells with PMA, using MALDI/MS combined with picomole-scale enzymatic treatment of proteolytic fragments. With this approach, low level constitutive phosphorylation at Ser^{25} and Ser^{36} was observed. However, phosphorylation at Ser^{25} was the major change in Op18a after treatment with PMA.

MATERIALS AND METHODS

Cell Culture and Metabolic ^{32}P Labeling—The human lymphoid leukemic T cell line, Jurkat, was cultured in RPMI 1640 medium (GIBCO/BRL) supplemented with 10% fetal calf serum. Prior to harvest, cells were treated with PMA (1 ng/ml) for 10 min. Cells were harvested by centrifugation at 1200 rpm for 5 min. The cell pellet was washed twice with phosphate-buffered saline. In some experiments, Jurkat cells were maintained in serum-free medium for 11 days at which time cell density reached a plateau of 1.8×10^6/ml for 4 days during which cell viability remained constant at 67%. For metabolic labeling, cells (10×10^6/ml) were suspended in phosphate-free RPMI 1640 containing [^{32}P]orthophosphate, carrier-free (Amersham Corp.; 400 μCi/ml), for 2 h at 37 °C. At the end of the labeling period, PMA (Sigma) was added to the cell suspension at 1 ng/ml for 10 min, and cells were harvested.

Purification of Op18 Phosphorylated Forms—The major Op18 phosphorylated forms, Op18a and Op18b, were purified by preparative two-dimensional PAGE using an IPG-based isoelectric separation in the first dimension as previously described (22). Cells were lysed with a buffer containing 8 M urea, 2% (v/v) Nonidet P-40, 0.8% (w/v) Ampholine, pH 3.5–10, 2% (v/v) mercaptoethanol, 0.9 mM phenylmethylsulfonyl fluoride, 25 mM NaF, and 2 mM sodium vanadate. IPG gels, pH 4–7 (17-cm separation distance, Pharmacia/LKB, Broma, Sweden), were rehydrated with 8 M urea, 2% Nonidet P-40, and 0.1 mM orange G as an indicator dye. Protein was applied to the anodes and focusing was for 100 kV-h. The second dimension SDS gels were an 11.5–14% acrylamide gradient. Proteins were electrophoretically transferred to nitrocellulose or PVDF (Immobilon P, Millipore, Bedford, MA) with a semidry blotting system (SemiPhor, Hoefer). Transfer buffer was 50 mM sodium borate containing either 20% (v/v) methanol (anode) or 5% (v/v) methanol (cathode). Electrophoretic transfer was at 1.3 mA/cm^2 for 2 h. ^{32}P-Labeled Op18a and Op18b were localized by autoradiography on Kodak XAR film and excised. Protein was eluted from PVDF membranes with 5% Tween 20 in 10 mM sodium phosphate, pH 7.0, for 4 h at 37 °C.[2] Protein was precipitated with 15% trichloroacetic acid overnight at 4 °C. Precipitated protein was collected by centrifugation and washed three times with acetone at −20 °C.

Proteolytic Digestion and Peptide Isolation—In situ trypsin digestion was performed as described by Aebersold (23). Briefly, nitrocellulose pieces were blocked with 0.5% (w/v) polyvinyl pyrrolidone-360 (Sigma) in 100 mM acetic acid for 30 min at 37 °C and washed extensively with deionized water and digestion buffer. Digestion was performed with modified trypsin (Promega, Madison, WI) (1:10, enzyme to substrate) in 50 mM ammonium bicarbonate at 37 °C. Following 4 h of incubation, a second aliquot of trypsin was added, and the incubation was continued for another 12 h. At the end of digestion, the solution and a buffer rinse were combined and dried in a Speed Vac (Savant, Farmingdale, NY). Alternatively tryptic digestion was performed on protein eluted from PVDF membranes as for *in situ* digestion. Tryptic peptides were redissolved in 0.1% (v/v) trifluoroacetic acid·H_2O (solvent A) and injected onto a C_{18} column (Vydac, 2.1 × 150 mm). Peptides were separated using a linear gradient from 5% to 60% solvent B (0.07% (v/v) trifluoroacetic acid in 80% (v/v) $CH_3CN \cdot H_2O$) in 55 min at a flow rate of 150 μl/min. Radioactivity of all the fractions were checked by Cerenkov counting. Radioactive fractions were used for thin layer phosphopeptide mapping and phosphorylation site studies. Thin layer phosphopeptide mapping was done as described by Boyle *et al.* (24). An aliquot of peptide was dried and redissolved in 5 μl of the first dimension buffer and spotted on a silica TLC plate (EM Separations, Gibbstown, NJ). Electrophoresis in the first dimension was at 1000 V for 60 min at 15 °C in acetic acid, formic acid (88%), water (156:50:1794) The plate was air-dried and transferred to a chromatography chamber which contained butanol-1:pyridine:acetic acid:water (785:607:122:486). Following ascending chromatography, the plate was air-dried and autoradiography was done using a PhosphorImager (Molecular Dynamics, Sunnyvale, CA).

N-terminal Sequencing—N-terminal sequencing was performed using a gas phase sequenator (model 473A, Applied Biosystems, Foster City, CA).

Secondary Enzymatic Digestion—Phosphopeptides were dephosphorylated with bacterial alkaline phosphatase. An aliquot of peptide (0.5–2 pmol) was dried in a Speed Vac and redissolved in 10 μl of 25 mM NH_4HCO_3 buffer, pH 7.9, and 1 μl of bacterial alkaline phosphatase (150 units, GIBCO/BRL). Incubation was at 37 °C for 1 h. The peptide was dried in a Speed Vac and reconstituted with 50% (v/v) CH_3CN in 0.1% (v/v) trifluoroacetic acid for MALDI/MS analysis. Digestion with chymotrypsin (Sigma) and endoproteinase Glu-C (Boehringer Manneheim) was performed on 0.5–2 pmol of phosphopeptide in 25 mM NH_4HCO_3 buffer, pH 7.9 (1:10 enzyme to substrate ratio). Incubations were at 37 °C for 18 h. They were then dried and reconstituted for MALDI/MS analysis.

Mass Spectrometry—MALDI mass spectra were obtained on a VT-2000 (Vestec Corp., Houston, TX) linear time-of-flight mass spectrometer equipped with a nitrogen laser (337 nm, 3-ns pulse). The accelerating voltage in the ion source was 30 kV. Data were acquired with a transient recorder with 5-ns resolution. Mass resolution ($m/\Delta m$) ranged from 200 to 400 (full width at half-maximum) depending on the sample and laser power. The matrix used for this study was α-cyano-4-hydroxycinnamic acid which was experimentally determined to yield the best results for the peptides analyzed. The matrix was dissolved in aqueous acetonitrile (33% v/v) to give a saturated solution at 20 °C (about 20 mM). To prepare a sample for analysis, 1 μl of the peptide solution (0.5–2 pmol/μl in 0.1% aqueous trifluoroacetic acid, acetonitrile, 1:1) was added to 1 μl of the matrix solution and applied to a stainless steel probe tip. The mixture was then allowed to air dry before being introduced into the mass spectrometer. Each spectrum was produced by accumulating 64 laser shots. Spectra were obtained from different regions of the probe tip, and a representative spectrum was selected for analysis. Time-to-mass conversion was done by internal calibration using peaks for Na^+ (m/z 22.99), K^+ (m/z 38.96), and matrix peaks at m/z 190.17 and 379.34. Potential peptide matches were made to the sequence of Op18 predicted from the cDNA (25) allowing for post-translational modification, *i.e.*, loss of N-terminal methionine and acetylation of alanine at residue 2 (13), using MacProMass software (S. Vemuri and T.D. Lee, Beckman Research Institute of the City of Hope, Duarte, CA).

RESULTS

Increased Metabolic Phosphorylation of Op18 by PMA Activation—We have recently shown that resting T cells express low but readily detectable levels of two major Op18 phosphorylated forms Op18a and Op18b and that activation of T cells through the T cell receptor complex in the presence of monocytes results in rapid Op18 phosphorylation (1). When the T cell receptor complex is bypassed by treatment of PBL with PMA, rapid phosphorylation of Op18 also occurs (26). Jurkat T cells which express high levels of Op18 were used in the present studies to characterize Op18 phosphorylation sites. Quiescent Jurkat cells were obtained as described under "Materials and Methods." Metabolic incorporation of [^{32}P] orthophosphate into growth-arrested Jurkat cells revealed the predominance of Op18a relative to other phosphorylated forms of Op18 (Fig. 1*A*). Other more acidic Op18 phosphoryl-

[2] J. Leykem and J. R. Strahler, manuscript in preparation.

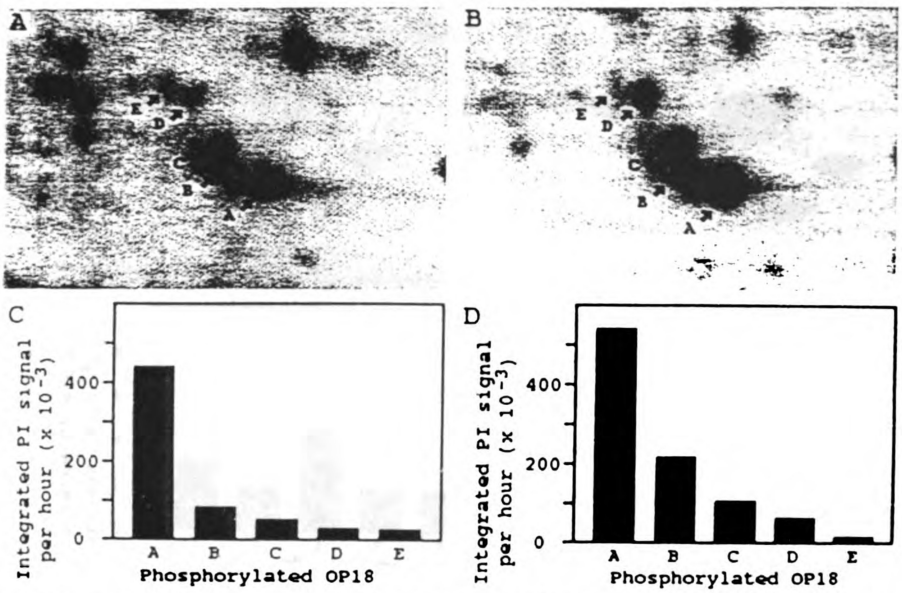


FIG. 1. **Two-dimensional PAGE of phosphorylated Op18.** Cells (3×10^6) were metabolically labeled as described under "Materials and Methods." Polypeptides were separated by two-dimensional PAGE and phosphorylated polypeptides visualized with a PhosphorImager. Shown is the region of each gel from 14 (*bottom*) to 23 kDa (*top*) and from pH 4.7 (*left*) to 6.5 (*right*). *A*, growth-arrested Jurkat T cells; *B*, PMA-treated Jurkat T cells; *C* and *D* are the quantitation results of the phosphorylated forms of Op18 obtained on PhosphorImager, for *panels A* and *B*, respectively.

ated forms b, c, d, and e (*spots B, C, D,* and *E* in Fig. 1) were readily detected. Minor forms c, d, and e were expressed at slightly elevated levels compared to the levels seen in growing cells. The majority of other phosphorylated polypeptides observed on two-dimensional gels was also qualitatively similar in growth-arrested cells as those seen in growing cells (data not shown). Following PMA treatment for 10 min there was increased incorporation of ^{32}P into all phosphorylated forms of Op18. The increment was greatest for Op18a and Op18b. More modest increments were observed for Op18c, d and e. The radioactivity in each form was determined by PhosphorImager analysis. In a representative experiment (Fig. 1), metabolic incorporation of ^{32}P into Op18a was 6-fold greater than that for Op18b in untreated Jurkat cells. Following 10 min of PMA treatment there was a 2–3-fold greater increase in phosphorylation of Op18b, c, d, and e compared to the increase in Op18a.

Identification of Phosphorylated Peptides of Op18 following PMA Activation—To identify peptide fragments containing the Op18 serine residues which are constitutively phosphorylated in Jurkat cells as well as those phosphorylated in response to PMA activation, [^{32}P]-labeled forms of Op18a and Op18b were isolated from preparative two-dimensional gels. Polypeptides were electrophoretically transferred to PVDF membranes, stained with Coomassie Blue, and protein-eluted from the membrane as described under "Materials and Methods." Following tryptic digestion, peptides from an aliquot of each digest were separated by two-dimensional thin layer peptide mapping. Phosphopeptides were detected using a PhosphorImager (Fig. 2). Five tryptic phosphopeptides were observed for Op18a isolated from growth-arrested Jurkat cells (Fig. 2). The same five phosphopeptides were observed for Op18a from PMA-activated Jurkat cells (Fig. 2) or from untreated growing Jurkat cells (data not shown). PMA activation resulted in a substantial increase in the radioactivity associated with peptide 3 (Fig. 2). Phosphopeptide mapping of tryptic digests of Op18b from growth-arrested and PMA treated cells revealed the same five radiolabeled tryptic fragments (Fig. 3). Tryptic peptides isolated from Op18b of PMA-activated cells all showed increased incorporation of radioactivity. The largest increment of incorporation was seen for peptides 1 and 3 and represented a 3-fold increase in incorporation.

For isolation of phosphopeptides corresponding to spots observed by thin layer chromatography a tryptic digest of Op18a derived from PMA-activated cells was separated by reverse-phase HPLC. Three major ^{32}P-labeled peptide fractions, designated TP1, TP2, and TP3, were obtained which accounted for greater than 80% of the total applied radioactivity. Fractions TP2 and TP3 contained the greatest amount of radioactivity.

To establish the identity of each reverse-phase HPLC fraction with the peptides of Fig. 2*B*, each of the three reverse-phase HPLC fractions were analyzed by two-dimensional thin layer peptide mapping individually and as a mixture with a trace amount of a total ^{32}P-tryptic digest of Op18a. TP3 contained a single phosphopeptide which co-migrated with peptide 3 (data not shown), and which increased by the greatest amount in response to PMA activation. Fraction TP1 contained peptide 1, while TP2 contained peptide 2 and a small amount of peptide 1. The two minor phosphopeptides observed on TLC were not recovered in sufficient amount for further characterization.

Partial Edman Sequence Identification of Phosphopeptides Recovered from Reverse-phase HPLC—Op18a was eluted from PVDF membranes of preparative, IPG-based two-dimensional gels and digested with trypsin as described under "Materials and Methods." Partial Edman sequence degradation was obtained for the phosphopeptides recovered from reverse-phase HPLC. Peptide TP1 had the sequence

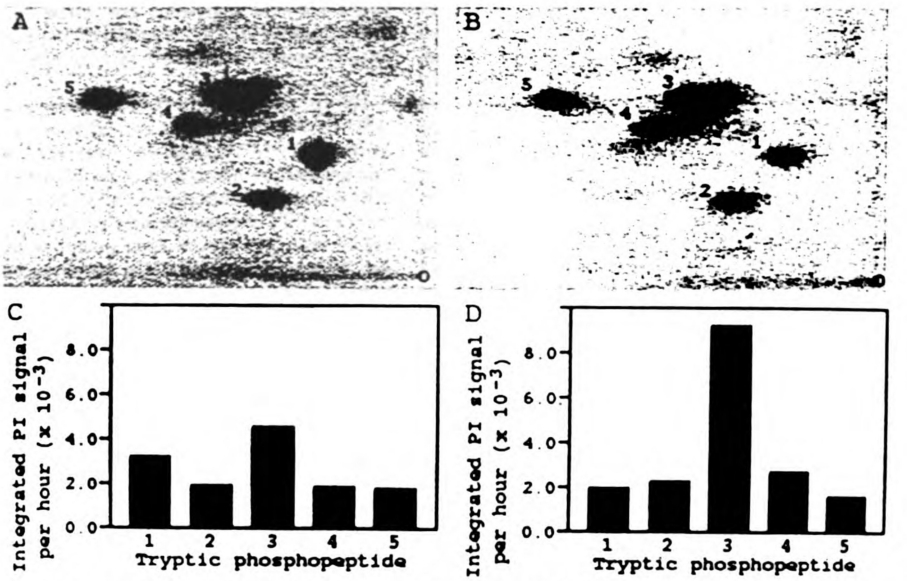


FIG. 2. **Two-dimensional thin layer phosphopeptide mapping of the tryptic phosphopeptides of Op18a.** Op18a was eluted from PVDF membranes and digested with trypsin as described under "Materials and Methods." *A*, tryptic phosphopeptides of Op18a from growth-arrested Jurkat T cells; *B*, tryptic phosphopeptides of Op18a from PMA-treated Jurkat T cells; *C* and *D* are the quantitation results obtained on PhosphorImager for *panels A* and *B*, respectively.

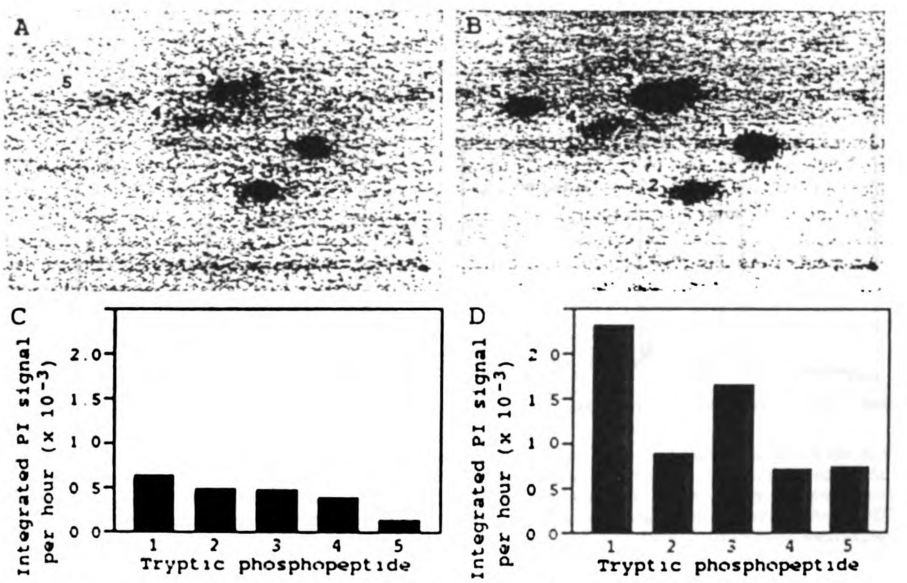


FIG. 3. **Two-dimensional thin layer phosphopeptide mapping of the tryptic phosphopeptides of Op18b.** Op18b tryptic phosphopeptides were prepared and analyzed as described in Fig. 2. *A*, tryptic phosphopeptides of Op18b from growth-arrested Jurkat T cells; *B*, tryptic phosphopeptides of Op18b from PMA-treated Jurkat T cells; *C* and *D* are the quantitation results obtained on PhosphorImager for *panels A* and *B*, respectively.

*XX*VP*X*FPLsP*X*[3] which agrees with the predicted tryptic peptide ESVPEFPLSPPK (residues 30–41) with two potential phosphorylation sites, Ser[31] and Ser[38]. For peptide TP2, the sequence *X*KEsVPEFPLs*X* was obtained. This sequence is consistent with an incomplete tryptic cleavage fragment, SKESVPEFPLSPPK (residues 28–41), containing potential phosphorylation sites at Ser[28], Ser[31], and Ser[38]. TP2 had a secondary sequence, *XX*VP*X*F, which represents contamination of TP2 with a small amount of TP1. TP3 gave the sequence *XXX*QAFEL*X*L*X* consistent with tryptic peptide ASGQAFELILSPR (residues 15–27) also with two potential phosphorylation sites, Ser[16] and Ser[25]. Limited amino acid

[3] Lower case letters are tentative calls. *X* is no call could be made either due to multiple residues or low signal to noise.

sequence information, which was consistent with the above results, was also obtained for these phosphopeptides derived from *in situ* trypsin digestion of Op18a electroblotted onto nitrocellulose.

MALDI/MS Characterization of Op18a Phosphopeptides—Further structural characterization of phosphopeptides TP1, TP2, and TP3 was undertaken by MALDI/MS to confirm the peptide assignments obtained by Edman sequencing and to distinguish between possible phosphorylation sites within each peptide. The spectrum of phosphopeptide TP3 is shown in Fig. 4*A*. The most abundant peak at *m/z* 1471 was within 2 mass units of the expected MH^+ (average mass) of two possible phosphorylated fragments (p2–13 and p15–27, *m/z* 1469.5 and 1469.6, respectively), or one nonphosphorylated incomplete tryptic fragment (101–112, *m/z* 1471.7) of Op18a. Because of the relatively poor mass resolution of a linear time-of-flight mass spectrometer and subsequent difficulty in assigning precise *m/z* values, it is not possible to distinguish among these three possibilities. The results of Edman sequencing indicated that the *m/z* 1471 peak was p15–27, phosphorylated at either Ser^{16} or Ser^{25}. Although the sequencing results indicated that the predominant peptide in TP3 was p15–27, p2–13 which is N-terminally blocked could conceivably be present in TP3. Three other peaks in the spectrum had *m/z* values consistent with their being nonphosphorylated tryptic peptides of Op18a (see Fig. 4 legend). Two of these

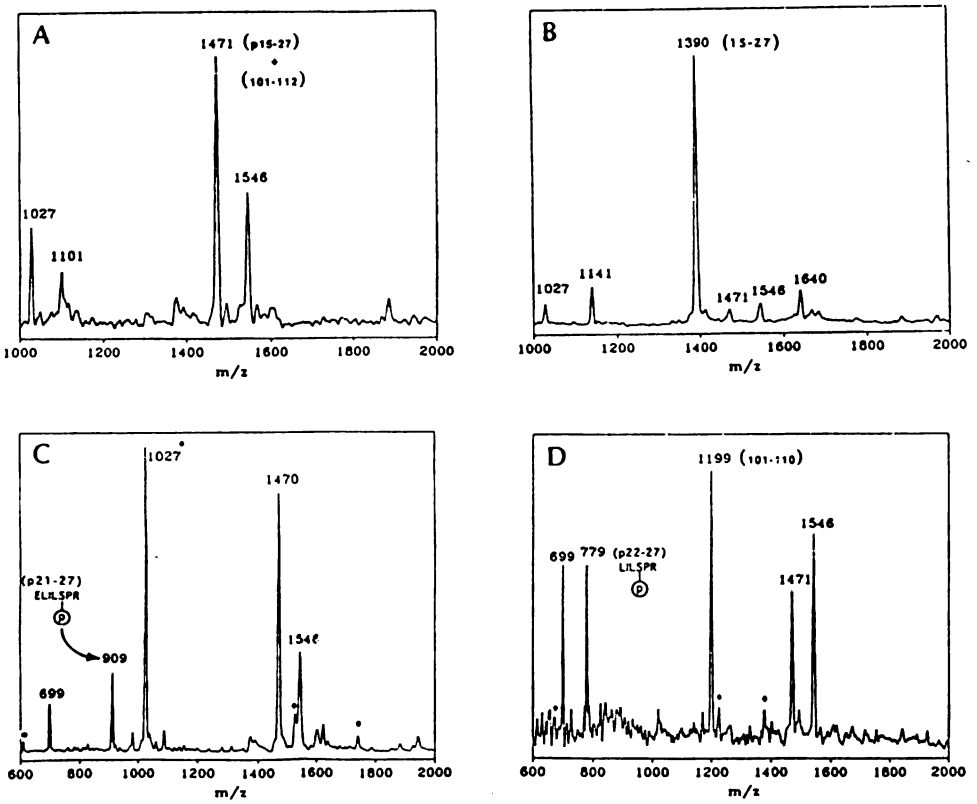


FIG. 4. **MALDI/MS of TP3.** Peptide assignments were made as described in the text. Other peaks are indicated in the MALDI spectra which were reproducibly observed and which could possibly result from complete or incomplete tryptic or secondary enzymatic digestion of Op18a phosphorylated at a single site. Asterisks (*) indicate background ions which were seen in the secondary enzyme control spectra. *A*, the spectrum of 2 pmol of TP3. Possible assignments for the peak at *m/z* 1471 include two phosphopeptides p2–13 (Ac-ASSDIQVKELEK having $pSer^2$ or $pSer^3$, theoretical 1469.5) and p15–27 (ASGQAFELILSPR having $pSer^{16}$ or $pSer^{25}$, 1469.6), and a nonphosphorylated fragment 101–112 (LTHKMEANKENR, 1471.7). The peak at *m/z* 1027 may represent tryptic peptide 127–134 (DKHIEEVR, 1026.1), while *m/z* 1101 could represent the incomplete tryptic digestion products 53–61 or 54–62 (KLEAAEERR and LEAAEERRK, respectively, both 1140.3). The ion at *m/z* 1546 may represent several possible tryptic products: 14–27 (RASGQAFELILSPR, 1545.8), 2–14 (Ac-ASSDIQVKELEKR, 1545.7), or 110–122 (ENREAQMAAKLER, 1546.8). *B*, the spectrum of TP3 (2 pmol) after alkaline phosphatase treatment displayed a new peak at *m/z* 1390, 81 atomic mass units below the peak at *m/z* 1471. Other possible tryptic fragment peaks appeared at *m/z* 1141 and 1640 possibly representing 77–85 (EHEKEVLQK, 1140.3) and 63–76 (SHEAEVLKQLAEKR, 1638.9), respectively. *C*, the spectrum of TP3 (2 pmol) following digestion with chymotrypsin. The peak at *m/z* 909 could be assigned to p21–27 (ELILpSPR, 908.0), with phosphorylation at Ser^{25}. The other prominent new peak at *m/z* 699 was not assignable to a chymotryptic fragment of any peptide detected in the previous spectra, but it may represent a tryptic peptide 76–80 (REHEK, 698.8). The increased intensity of the signal at *m/z* 1027 probably resulted from an overlapping background peak found in the chymotrypsin control spectrum. *D*, the spectrum of TP3 (2 pmol) following endoproteinase Glu-C digestion. The peak at *m/z* 779 represents the endoproteinase Glu-C fragment p22–27 (LILpSPR, 779.8) of p15–27, indicating again that Ser^{25} is phosphorylated. The ion at *m/z* 699 might be due to peptide 76–80 seen in the TP3 chymotrypsin spectrum (*C*), but alternatively could be derived from digestion of the nonphosphorylated peptide ion at *m/z* 1546 if the latter peak represented 14–27. This would yield an endoproteinase Glu-C fragment at *m/z* 698 (22–27). This same fragment could also be formed by dephosphorylation of p22–27. The peak at *m/z* 1199 could represent 101–110 (LTHKMEANKE, 1201.4); a partial endoproteinase Glu-C digestion product if *m/z* 1471 represented 101–112.

(*m/z* 1027 and 1101) were observed at lower intensities in MALDI spectra of other peptide fractions (*e.g.*, Fig. 5*A*).

The relative intensity of signals in the MALDI mass spectrum do not necessarily correlate with the relative concentrations of different peptides in the sample. At the low picomole level, the peptide-matrix ratio and other factors, such as the presence of other peptides and salts, may result in an increase or diminution of individual peptide signals. All significant ions were considered throughout this and subsequent analyses as potentially representing peptides derived from the appropriate digestions of Op18 and its tryptic peptides. In some instances secondary digestion products were obtained which provided evidence for the tentative peptide assignment given to these peaks.

To confirm the identity of *m/z* 1471 as a phosphopeptide, an aliquot of TP3 was treated with alkaline phosphatase and analyzed by MALDI/MS. The new spectrum showed that the peak at *m/z* 1471 shifted to *m/z* 1390 (−80 mass units), in accord with the removal of a phosphate group (Fig. 4*B*). The other peaks in the original spectrum did not shift. Although the peak at *m/z* 1546 decreased, no phosphorylated peptide could be predicted from the sequence of Op18a with this mass and no new peak appeared at *m/z* 1466 expected for a dephosphorylated product. Relative signal intensity also varied for peaks at *m/z* 1141 (*cf.* Fig. 4, *A* and *B*) and m/z 1640 (*cf.* Fig. 4, *B* and *C*). These peaks could represent other nonphosphorylated partial tryptic fragments present in TP3 (see Fig. 4 legend).

p15–27 contains chymotryptic and endoproteinase Glu-C cleavage sites, Phe^{20} and Glu^{21}, respectively, located between possible phosphorylation sites, while p2–13 contains endoproteinase Glu-C cleavage sites Glu^{10} and Glu^{12}. To distinguish between these two peptides and to determine the phosphorylation site, an aliquot of TP3 was treated with chymotrypsin and analyzed by MALDI/MS (Fig. 4*C*). The peak at *m/z* 909 was consistent with a phosphorylated chymotryptic fragment, ELILpSPR (p21–27, theoretical *m/z* 908.0), indicating that Ser^{25} was phosphorylated. The increased signal intensity of the *m/z* 1027 peak likely results from an overlapping background peak found in the chymotrypsin control spectrum (data not shown). TP3 digested with endoproteinase Glu-C (Fig. 4*D*) showed a signal at *m/z* 779 expected for p22–27 (LILpSPR, theoretical *m/z* 779) confirming Ser^{25} as the phosphorylation site in p15–27. No peaks were observed that were consistent with expected endoproteinase Glu-C fragments of p2–13. The peak at *m/z* 1199, however, could be assigned to 101–110 (LTHKMEANKE, *m/z* 1201.4), a partial endoproteinase Glu-C fragment of 101–112, indicating that *m/z* 1471 represented a mixture of p15–27 and 101–112. The resolution of our MALDI time of flight mass spectrometer did not permit these two peptides, which differ by 2 atomic mass units, to be resolved. The peak at *m/z* 699 could be an endoproteinase Glu-C fragment of *m/z* 1546 or dephosphorylated p22–27; however, this ion was also present in the chymotryptic digest. Taking the results of partial Edman sequencing and MALDI/MS analysis of secondary enzymatic cleavage products, we conclude that *m/z* 1471 in the MALDI spectrum corresponds to peptide p15–27, ASGQAFELILpSPR, phosphorylated at Ser^{25}.

MALDI/MS analysis of phosphorylated TP2 (Fig. 5*A*) contained a peak at *m/z* 1627, which could represent four possible phosphorylated tryptic fragments of Op18a (p2–14, *m/z* 1625.7; p14–27, *m/z* 1625.8; p28–41, *m/z* 1622.8; and p136–149, *m/z* 1629.6). Assignment of *m/z* 1627 as p28–41 is consistent with the primary amino acid sequence obtained by Edman degradation. The predominant peak at *m/z* 1330 is consistent with the secondary Edman sequence obtained for TP2. The relative abundance of these peptides could vary in MALDI spectra as discussed above or could be due to a difference in initial coupling yield or peptide retention during Edman degradation. Other peaks in the spectrum could represent nonphosphorylated tryptic peptides (see Fig. 5 legend). Partial dephosphorylation subsequent to HPLC isolation may have produced the peak at *m/z* 1550; however, contribution from other peptides is possible. Alkaline phosphatase treatment eliminated the peak at *m/z* 1627 and shifted the center of mass of the signal at *m/z* 1550 to 1546, in accord with the removal of a phosphate group from the peptide at *m/z* 1627 (Fig. 5*B*).

To further confirm identification of *m/z* 1627 as p28–41 and to identify the phosphorylation site, secondary digestion was performed with chymotrypsin and endoproteinase Glu-C. TP2 was resistant to significant degradation by chymotrypsin (Fig. 5*C*). Two new peaks at *m/z* 915 and 785 could not be assigned as digestion products of p28–41 nor of any other possible phosphorylated or nonphosphorylated tryptic peptides. The *m/z* 785 peak was also seen after endoproteinase Glu-C digestion. Both peaks could represent tryptic digestion products of Op18 (see Fig. 5 legend). A minor peak at *m/z* 716 fit the expected mass for chymotryptic fragment p36–41 (PLpSPPK, theoretical *m/z* 718) of p28–41. None of the peaks having a signal intensity similar to that of *m/z* 716 could be assigned as digestion products. Digestion with endoproteinase Glu-C gave added support to the assignment of *m/z* 1627 in TP2 as p28–41. Although the spectrum (Fig. 5*D*) again indicated the digestion was not complete, a minor peak present at *m/z* 1280 was consistent with the mass for p30–41 (SVPEFPLSPPK phosphorylated at either Ser^{31} or Ser^{36}, theoretical *m/z* 1278). The resistance of TP2 (and TP1, see below) to significant digestion by chymotrypsin and endoproteinase Glu-C could be due to adjacent proline residues (Pro-Glu-Phe-Pro) surrounding these cleavage sites. The peak at *m/z* 785 could be an endoproteinase Glu-C product 35–41 of *m/z* 1330 peak (30–41), however the *m/z* 785 peak was also seen in the chymotryptic digest. None of the other minor peaks could be matched to any predicted digestion products. Thus assignment of the phosphorylation site in p28–41 could only be made based on the *m/z* 716 peak in the chymotryptic digest. A peak of similar intensity was observed in spectra taken from other regions of the probe tip and thus did not appear to be a spurious signal. We conclude that TP2 contained phosphorylated p28–41 with phosphorylation at Ser^{36}.

The MALDI/MS spectrum of TP1 displayed a weak signal at *m/z* 1409 (Fig. 6*A*). The peak shifted to *m/z* 1328 upon treatment with alkaline phosphatase confirming that *m/z* 1409 was a phosphopeptide (Fig. 6*B*). Three possible phosphorylated tryptic peptides of Op18 have an approximate mass of *m/z* 1409: p30–41 (*m/z* 1407.5), and p42–52 and p43–53 (both *m/z* 1411.6). Based on Edman sequencing results we assign *m/z* 1409 as p30–41. The MALDI/MS spectrum after chymotrypsin and endoproteinase Glu-C digestion indicated that the potential cleavage sites in TP1, which are the same sites as in TP2, were resistant to extensive digestion (data not shown). The spectra contained numerous low level background peaks from the respective enzyme reaction mixtures. Potentially informative fragments were at the level of these background peaks and no phosphorylation site could be positively identified.

DISCUSSION

These studies identify two major sites of phosphorylation in Op18a in PMA-activated Jurkat cells. Of five tryptic phos-

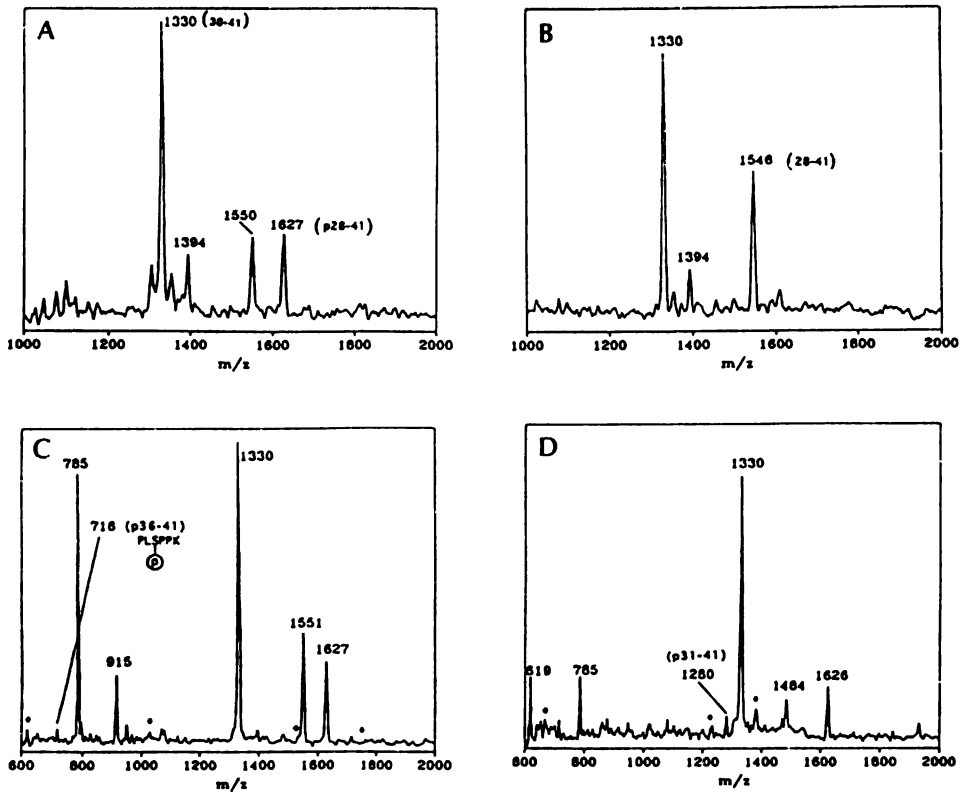


FIG. 5. MALDI/MS of TP2. *A*, the spectrum of 1 pmol of TP2. Potential phosphorylated tryptic peptides of Op18a in the TP2 spectrum all have *m/z* 1627. Candidates are p2–14 (Ac-ASSDIQVKELEKR having pSer[2] or pSer[3], theoretical 1625.7), p14–27 (RASGQAFELILSPR having pSer[16] or pSer[25], 1625.8), p28–41 (SKESVPEFPLSPPK having pSer[25], or pSer[31] or pSer[38], 1622.8), and p136–149 (NKESKDPADETEAD having pSer[139] or pThr[144], 1629.6). Prominent peaks which are possible tryptic fragments include: *m/z* 1330, 30–41 (ESVPEFPLSPPK, 1327.5), 42–52 (KKDLSLEEIQK, 1331.6), or 43–53 (KDLSLEEIQKK, 1331.6); *m/z* 1550, 110–122 (ENREAQMAAKLER, 1546.8), or 136–149 (NKESKDPADETEAD, 1549.5); and *m/z* 1394, 127–137 (KHIEEVRKNK, 1396.6). *B*, the spectrum of TP2 (1 pmol) following alkaline phosphatase treatment. The peak at *m/z* 1627 shifted to *m/z* 1546 (81 atomic mass units lower). *C*, chymotryptic digestion of TP2 (1 pmol) did not produce a significant reduction of the peak at *m/z* 1627. However, the small peak at *m/z* 716 was consistent with the chymotryptic fragment, p36–41 (718.8) of tryptic peptide p28–41. The peak at *m/z* 785 may represent tryptic fragment 123–128 (REKDK, 788.9) or 129–134 (HIEEVR, 782.9), and the peak at *m/z* 915 may indicate tryptic peptide 63–70 (SHEAEVLK, 913.0) was present. *D*, the endoproteinase Glu-C spectrum of TP2 (1 pmol) also indicated incomplete digestion of the phosphopeptide. The peak at *m/z* 785 in the spectrum might be assigned to the endoproteinase Glu-C fragment 35–41 (FPLSPPK, theoretical 786.0) derived from the nonphosphorylated peptide 28–41 (*m/z* 1330), although it could be due to the tryptic fragment seen in the previous spectrum (*C*). The signal at *m/z* 619 could be assigned to the fragment 114–119 (AQMAAK, 619.8), while the peak at *m/z* 1484 might represent tryptic fragment 63–75 (SHEAEVLKQLAEK, 1482.7).

phopeptides detected by two-dimensional thin layer peptide mapping, one contained phosphorylated Ser[25] and two had phosphorylated Ser[38]. Two other minor tryptic phosphopeptides were not recovered by reverse-phase HPLC in amounts sufficient to perform characterization studies. These minor phosphopeptides may represent incomplete digestion with trypsin. However, we cannot exclude the possibility that they represent low levels of phosphorylation at sites other than Ser[25] and Ser[38] in Op18a and Op18b. We have observed other minor phosphorylated forms at specific stages of the cell cycle. Other phosphorylation sites must be utilized to generate these forms. Op18c has an isoelectric point similar to Op18b and likely contains a third site in addition to either Ser[25] or Ser[38]. Similarly Op18d and Op18e are more acidic and are presumed to contain additional phosphorylation sites. Phosphorylation at these additional sites must result in significant conformational alteration of the SDS-denatured protein since these forms display slightly increased apparent molecular weights in the second dimension.

Comparison of two-dimensional thin layer peptide maps revealed the same five tryptic phosphopeptides in Op18a of growth-arrested Jurkat cells as were seen in Op18a following PMA treatment. Based on quantitation of tryptic phosphopeptides of Op18a we conclude that constitutive phosphorylation in Op18 occurs at Ser[25] and at Ser[38] with a slight preference for phosphorylation at Ser[25] to generate Op18a. In response to treatment with PMA there is a large increment in phosphorylation of Ser[25]. The level of phosphorylation at Ser[38] is actually reduced somewhat.

Tryptic phosphopeptide maps of Op18b again reveal the same five phosphopeptides as are seen for Op18a indicating that phosphorylation to form Op18b occurs at the same sites as for Op18a. Since Op18b has an isoelectric point more acidic than Op18a by an amount consistent with one more negative charge, we conclude that Op18b is phosphorylated at both Ser[25] and Ser[38]. From the quantitation of tryptic phosphopeptides of Op18b, it appears that PMA treatment results in a significant increment in phosphorylation at Ser[38]. The results

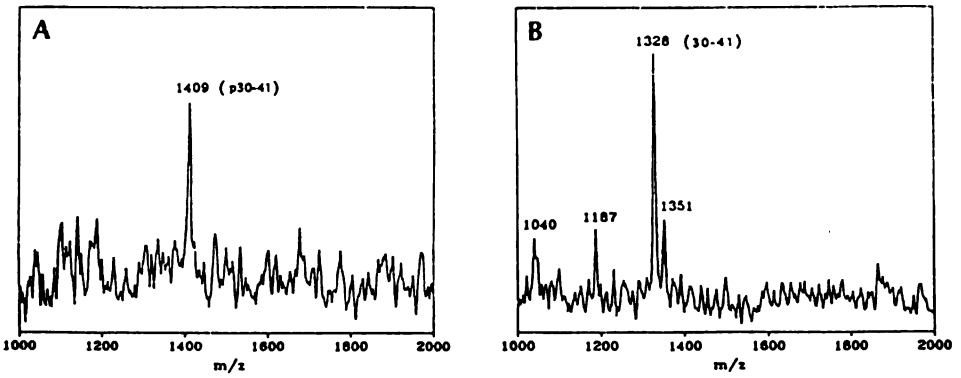


FIG. 6. **MALDI/MS of TP1.** *A*, the spectrum of 0.5 pmol of TP1. The observed peak at *m/z* 1409 could represent several phosphorylated tryptic peptides, p30–41 (ESVPEFPLSPPKK, theoretical 1406.7), p42–52 (KKDLSLEEIQK, 1411.6), or p43–53 (KDLSLEEIQKK, 1411.6), or a nonphosphorylated peptide 125–135 (EKDKHIEEVRK, 1411.6). *B*, the spectrum of TP1 (0.5 pmol) after alkaline phosphatase treatment. Alkaline phosphatase treatment produced a shift to *m/z* 1328 (−81 atomic mass units). The peak at *m/z* 1351 probably represents a sodium ion adduct of the *m/z* 1328 peak. Peaks at *m/z* 1040 and 1187 were present in both spectra and may represent tryptic peptides 62–70 (KSHEAEVLK, 1041.2) and 120–128 (LERLREKDK, 1187.4), respectively.

suggest that Op18 phosphorylation occurs in obligate order in response to PMA.

Several possibilities may explain this differential degree of phosphorylation at Ser^{38} in Op18a compared to Op18b. In one scenario, in unstimulated cells the sequence surrounding both Ser^{25} and Ser^{38} presents these residues as suitable substrates for a constitutively active protein kinase. Treatment with PMA results in activation of a second protein kinase which is selective for Ser^{25} of Op18 as the protein substrate. Phosphorylation at Ser^{25} may produce a conformational change in this region of the protein resulting in Ser^{38} becoming a suitable substrate for the second protein kinase. Another possibility is that while both Op18a and b are associated with a nonmembranous, soluble fraction, Op18a could be constitutively expressed in a subcellular compartment distinct from a subcellular compartment for Op18b. Incorporation of ^{32}P into Op18a occurs during the constitutive labeling period prior to PMA treatment in both compartments. PMA could then activate a protein kinase which recognizes the motif surrounding Ser^{25} as substrate thus converting a substantial portion of Op18a phosphorylated at Ser^{38} to doubly phosphorylated Op18b.

PMA exerts its primary effect by activating protein kinase C which results in a cascade of phosphorylation events including downstream activation of other protein kinases and phosphatases. The amino acid sequences surrounding Ser^{25} and Ser^{38} share several common features including the Ser-*X*-*X*-Arg/Lys or Ser-*X*-*X*-*X*-Arg/Lys sequence motif for protein kinase C substrates. Other evidence has suggested that protein kinase C may be involved in Op18 phosphorylation. Gullberg *et al.* (12) have shown that the protein isolated from human T cells can be phosphorylated by partially purified protein kinase C. Calphostin C, an inhibitor of protein kinase C, prevents the rapid increase in Op18 phosphorylation which occurs within 2 min in PBL following activation of the T cell receptor complex with OKT3 (1) or in human leukemic cells treated with PMA and Jurkat T cells treated with OKT3 (data not shown). Studies are in progress to determine which sites are phosphorylated in response to T cell receptor activation and in an *in vitro* phosphorylation system.

Since both Ser^{25} and Ser^{38} are followed by a Pro, the Ser-Pro motif prompted Labdon *et al.* (14) to suggest that Ser^{25} and Ser^{38} are proline-directed phosphorylation sites. They suggest that potential kinases include $p34^{cdc2}$, a Pro-directed Ser/Thr protein kinase and microtubule-associated protein kinase. $p34^{cdc2}$ is the catalytic subunit of a nuclear complex regulated during the cell cycle (27, 28). While $p34^{cdc2}$ does not have a nuclear localization in the absence of expression of cdc13, another member of the complex (27), there is no evidence for an active kinase complex involving cdc2 in the cytosolic compartment in which Op18 is located.

It is interesting that human Op18 in Jurkat cells is phosphorylated at the same two major sites as the protein counterpart, p19, in bovine brain (14). Based on their cDNA sequences the two proteins differ by only two residues of 149 amino acids. Also this extremely high degree of amino acid conservation exists between the human (25) and rat protein (29, 30). A single amino acid difference exists between the two species. Op18 phosphorylation may represent a common element in a number of signaling pathways. Phosphorylation of the counterpart of Op18a and Op18b in other species is the most prominent feature of signal transduction in response to diverse stimuli. The relative amount of other minor forms of phosphorylated Op18 changes as cells progress through the cell cycle (7). Op18c levels are highest during the transition from G_0 to G_1 whereas Op18d and Op18e are observed in M-phase arrested cells.

We have demonstrated that MALDI/MS in combination with enzymatic digestion can be a useful tool to obtain structural information of phosphopeptides especially in the cases where extremely low sample quantities are available. Each enzymatic treatment and MALDI spectrum in this study consumed from 0.5–2 pmol of sample. Other mass spectral techniques, such as electrospray MS, can reach this level of sensitivity, but they are not compatible with the buffer concentrations required for the enzymatic reactions. Although we have worked with tryptic fragments, in principle this approach could be applied to much larger polypeptides.[4] One factor that makes multiple enzymatic reactions necessary with MALDI/MS is the relatively poor mass resolution. Since many peptides derived from a single protein can yield ions close or identical in mass, some method must be applied to distinguish among the various possible combinations. Specific chemical reactions may be useful in this regard, provided they are efficient at the low picomole level.

[4] P. C. Andrews, unpublished data.

REFERENCES

1. Strahler, J. R., Hailat, N., Lamb, B. J., Rogers, K. P., Underhill, J., Melhem, R. F., Keim, D. R., Zhu, X.-X., Kuick, R. D., Fox, D. A., and Hanash, S. M. (1992) *J. Immunol.* **149**, 1191–1198
2. Doye, V., Boutterin, M.-C., and Sobel, A. (1990) *J. Biol. Chem.* **265**, 11650–11655
3. Sobel, A., Boutterin, M.-C., Beretta, L., Chneiweiss, H., Doye, V., and Peyro-Saint-Paul, H. (1989) *J. Biol. Chem.* **264**, 3765–3772
4. Mary, D., Peyron, J. F., Auberger, P., Aussel, C., and Fehlman, M. (1989) *J. Biol. Chem.* **264**, 14498–14502
5. Schubart, U. K. (1982) *J. Biol. Chem.* **257**, 12231–12238
6. Peyron, J. F., Aussel, C., Ferrua, B., Haring, H., and Fehlman, M. (1989) *Biochem. J.* **258**, 505–510
7. Strahler, J. R., Lamb, B. J., Ungar, D. R., Fox, D. A., and Hanash, S. M. (1992) *Biochem. Biophys. Res. Commun.* **185**, 197–203
8. Melhem, R. F., Strahler, J. R., Hailat, N., Zhu, X. X., and Hanash, S. M. (1991) *Biochem. Biophys. Res. Commun.* **179**, 1649–1655
9. Cooper, H. L., Fuldner, R., McDuffie, E., and Braverman, R. L. (1990) *J. Immunol.* **145**, 1205–1213
10. Hailat, N., Strahler, J., Melhem, R., Zhu, X. X., Brodeur, G., Seeger, R. C., Reynolds, C. P., and Hanash, S. M. (1990) *Oncogene* **5**, 1615–1618
11. Hanash, S. M., Strahler, J. R., Kuick, R., Chu, E. H. Y., and Nichols, D. (1988) *J. Biol. Chem.* **263**, 12813–12815
12. Gullberg, M., Noreus, K., Brattsand, G., Friedrich, B., and Shingler, V. (1990) *J. Biol. Chem.* **265**, 17499–17505
13. Hanash, S. M., Strahler, J. R., Neel, J. V., Hailat, N., Melhem, R., Keim, D., Zhu, X. X., Wagner, D., Gage, D. A., and Watson, J. T. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 5709–5713
14. Labdon, J. E., Nieves, E., and Schubart, U. K. (1992) *J. Biol. Chem.* **267**, 3506–3513
15. Hillenkamp, F., Karas, M., Beavis, R. C., and Chait, B. T. (1991) *Anal. Chem.* **63**, 1193A–1203A
16. Karas, M., and Bahr, U. (1991) *Mass Spectrom. Rev.* **10**, 335–357
17. Chait, B. T., and Kent, S. B. H. (1992) *Science* **257**, 1885–1894
18. Chait, B. T., Wang, R., Beavis, R. C., and Kent, B. H. (1992) in *Proceedings of the 40th ASMS Conference on Mass Spectrometry and Allied Topics, Washington, D. C.*, pp. 1939–1940, North American Society of Mass Spectrometry, Sante Fe
19. Schar, M., Bornsen, K. O., and Gassmann, E. (1991) *Rapid Commun. Mass Spectrom.* **5**, 319–326
20. Schar, M., Bornsen, K. O., Gassmann, E., and Widmer, H. W. (1991) *Chimia* **45**, 123–126
21. Spengler, B., Kirsch, D., Kaufmann, R., and Jaeger, E. (1992) *Rapid Commun. Mass Spectrom.* **6**, 105–108
22. Strahler, J. R., and Hanash, S. M. (1991) *Methods: A Companion to Methods in Enzymology* **3**, 109–111
23. Aebersold, R., Leavitt, J., Saavedra, R. A., and Hood, L. E. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 6970–6974
24. Boyle, W. J., van der Geer, P., and Hunter, T. (1991) *Methods Enzymol.* **201**, 110–149
25. Zhu, X. X., Kozarsky, K., Strahler, J. R., Eckerskorn, C., Lottspeich, F., Melhem, R., Lowe, J., Fox, D. A., Hanash, S. M., and Atweh, G. F. (1989) *J. Biol. Chem.* **264**, 14556–14560
26. le Gouvello, S., Chneiweiss, H., Tarantino, N., Debre, P., and Sobel, A. (1991) *FEBS Lett.* **287**, 80–84
27. Booher, R. N., Alfa, C. E., Hyams, J. S., and Beach, D. H. (1989) *Cell* **58**, 485–497
28. Riabowol, K., Draetta, G., Brizuela, L., Vandre, D., and Beach, D. (1989) *Cell* **57**, 393–401
29. Doye, V., Soubrier, F., Bauw, G., Boutterin, M.-C., Beretta, L., Koppel, J., Vandekerckhove, J., and Sobel, A. (1989) *J. Biol. Chem.* **264**, 12134–12137
30. Schubart, U. K., Banerjee, M. D., and Eng, J. (1989) *DNA* **8**, 389–398

Appendix VI - Source code of MSU MassMap v.1.1, a mass mapping program for phosphoprotein analysis

```
'*** <<< this is used to precede all comments

'*** general: declare variables and constants
  '*** AA$(3000): array of amino acid sequence; CutAA_$(50): amino acid residues to be cut;
  '*** CutCN$(50): cut N-terminal or C-terminal
  Dim AA$(3000), CutAA_$(50), CutCN$(50)
  '*** CutPos(500): define cut position,    (n-1) refer to the N-terminal of n-th residue,
  '***                                      n refer to the C-terminal of n-th residue;
  '*** FrL1(30000): starting (left) residue number of a fragment;
  '*** FrR1(30000): ending (right) residue number of a fragment;
  '*** the following two arrays consider only those fragments within mass range;
  '*** FrL2(30000); FrR2(30000);
  '*** pNum(30000): number of phosphorylation sites in that fragment;
  '*** MsIdx(30000): index array for the sorting by mass value;
  '*** LSrchMtch(800): starting (left) residue number of a fragment which meets mass search;
  '*** RSrchMtch(800): ending (right) residue number of a fragment which meets mass search
  Dim CutPos(500) As Integer, FrL1(30000) As Integer, FrR1(30000) As Integer, FrL2(30000) As
  Integer, FrR2(30000) As Integer, pNum(30000) As Integer, MsIdx(30000) As Integer, LSrchMtch(800)
  As Integer, RSrchMtch(800) As Integer
  '*** AmsAA(3000): average mass of an amino acid residue;
  '*** FrMs(30000): mass of a fragment;
  '*** MsSrchMtch(800): mass of a fragment which meets mass search requirement
  Dim AmsAA(3000), FrMs(30000), MsSrchMtch(800)
  '*** peptide_name$: name of the protein/peptide; FileName$: name of file contains protein sequence;
  '*** FileName2$: name of file to save calculation result output;
  '*** peptide$: a temporary variable to keep amino acid residues read from protein sequence file;
  '*** CutCorN$: cut C-terminal or N-terminal; TempStr$: temporary string;
  '*** pSer$, pThr$, pTyr$: a switch to check if the phosphorylation on that residue is considered;
  '*** SeqnSrch$: the sequence to be searched for
  Dim peptide_name$, FileName$, FileName2$, peptide$, CutCorN$, pSer$, pThr$, pTyr$, TempStr$,
  SeqnSrch$
  '*** CAANum: number of kinds of amino acid to be cut;
  '*** CPos:  a temporary store of cut position, (n-1) refer to the N-terminal of n-th residue,
  '***                                      n refer to the C-terminal of n-th residue;
  '*** CutNum: total number of cuts; NumberAA: total number of amino acid residues
  Dim CAANum As Integer, CPos As Integer, CutNum As Integer, NumberAA As Integer
  '*** FrNum1: number of fragments; FrNum2: number of fragments within mass range;
  '*** pFrNum: number of fragments which are phosphorylated;
  '*** pNumTemp: temporary storage of number of fragments which are phosphorylated
  Dim FrNum1 As Integer, FrNum2 As Integer, pFrNum As Integer, pNumTemp As Integer
  '*** CountL, CountR: temporary counts for mass search routine, to keep left/ right of temporary
  peptides;
```

```
'*** CountMtch: count of number of matches
Dim CountL As Integer, CountR As Integer, CountMtch As Integer
'*** PepMW: molecular weight of whole protein/peptide;
'*** AmsAAfirst, AmsAAlast: average mass of first/last amino acid residue;
'*** MsLowLimit, MsHighLimit: low/high limit of mass range; Temp: temporary storage;
'*** Nmodify, Cmodify: mass shift due to N-terminal or C-terminal modification;
'*** ValSrchMs, RngSrchMs: mass value/error range for the mass search
Dim PepMW, AmsAAfirst, AmsAAlast, MsLowLimit, MsHighLimit, Temp, Nmodify, Cmodify,
ValSrchMs, RngSrchMs
'*** SrchMsTemp: a double-precision temporary used in mass search routine
Dim SrchMsTemp As Double
'*** assign mass to residues using constants
Const AMsA = 71.0788, AMsR = 156.1875, AMsN = 114.1038, AMsD = 115.0886, AMsC =
103.1388, AMsE = 129.1155, AMsQ = 128.1307, AMsG = 57.0519, AMsH = 137.1411, AMsI =
113.1594
Const AMsL = 113.1594, AMsK = 128.1741, AMsM = 131.1926, AMsO = 114.1472, AMsF =
147.1766, AMsP = 97.1167, AMsS = 87.0782, AMsT = 101.1051, AMsW = 186.2132, AMsY =
163.176, AMsV = 99.1326
'*** constants for shortcuts
Const EIGHTY = 79.97986, NINETEEN = 19.0231

'*** pull down menu of Help: about this program
Sub about_Click ()
 Beep: Form1.Cls: For I = 1 To 20: Form1.Print : Next I
 Form1.Print "        This program was written by Pao-Chi Liao at Mass"
 Form1.Print "  Spectrometry Facility, Biochemistry, Michigan State Uni-"
 Form1.Print "  versity, to calculate the average masses of possible pep-"
 Form1.Print "  tide fragments and phosphopeptide fragments from specific"
 Form1.Print "  enzymatic or chemical degradation of a protein and the"
 Form1.Print "  m/z value of the mass spectral peak for the corresponding"
 Form1.Print "  [M+H]+ ion.  In this program, 'degradation' merely means"
 Form1.Print "  hydrolysis. The average masses were used instead of mono-"
 Form1.Print "  isotopic masses to make the program suitable for the mass"
 Form1.Print "  mapping analyses for MALDI-TOF-MS measurements of pro-"
 Form1.Print "  teins or peptides. The program computes masses for pep-"
 Form1.Print "  tides from not only complete but also partial digestion."
 Form1.Print "  For a detailed discussion on how to locate phosphorylation"
 Form1.Print "  sites in a phosphoprotein using MALDI-TOF-MS and mass "
 Form1.Print "  mapping, see P.-C. Liao, J. Leykam, P. Andrews, D. Gage"
 Form1.Print "  and J. Allison, ANALYTICAL BIOCHEMISTRY, 219, 9-20 (1994)."
 MsgBox "", 0, ""
 Form1.Cls
End Sub

'*** after buttom: cleavage at C-terminal
Sub After_Click ()
 Beep
 If After.Value Then
  CutCorN$ = "After"
 End If
End Sub

'*** before buttom: cleavage at N-terminal
Sub Before_Click ()
```

```
 Beep
 If Before.Value Then
  CutCorN$ = "Before"
 End If
End Sub

'*** inputbox: define C-terminal modification by mass
Sub C_modify_Change ()
 Beep
 Cmodify = Val(C_modify.Text)
End Sub

'*** pull down menu: initiate calculation of fragments
Sub CalFrag_Click ()
 Beep
 FrNum1 = 0: FrNum2 = 0: pFrNum = 0: CutNum = 0
 '***enable items in menu
 Drive1.Enabled = 0: Dir1.Enabled = 0: File1.Enabled = 0: : Drive1.Visible = 0: Dir1.Visible = 0:
File1.Visible = 0
 PrinterCalFrag.Enabled = 1: FileCalFrag.Enabled = 1: ScreenCalFrag.Enabled = 1
 Label12.Visible = 1
 '***modify the mass of both termini for modification
 AmsAA(1) = AmsAAfirst + Nmodify: AmsAA(NumberAA) = AmsAAlast + Cmodify
 '***find cut positions
 Form1.Cls
 For I = 1 To 13
  Form1.Print
 Next I
 For I = 1 To NumberAA
  For J = 1 To CAANum
   If AA$(I) = CutAA_$(J) Then
    If CutCN$(J) = "Before" Then
     CPos = I - 1
    Else
     CPos = I
    End If
    CutNum = CutNum + 1
    CutPos(CutNum) = CPos
   End If
  Next J
 Next I
 If CutPos(1) <> 0 Then
  CutNum = CutNum + 1
  For I = CutNum To 2 Step -1
   CutPos(I) = CutPos(I - 1)
  Next I
  CutPos(1) = 0
 End If
 If CutPos(CutNum) <> NumberAA Then
  CutNum = CutNum + 1
  CutPos(CutNum) = NumberAA
 End If
 Form1.Print " Protein/Peptide : "; peptide_name$
 Form1.Print " cut positions are:"
```

```
For I = 2 To CutNum - 1
 Form1.Print CutPos(I);
 If (I - 1 - Int((I - 1) / 10) * 10) = 0 Then
  Form1.Print
 End If
Next I
Form1.Print
If CutNum = 1 Then
  Form1.Print " total # of cuts= "; " 0"
Else
 Form1.Print " total # of cuts= "; CutNum - 2
End If

'***find fragments by number pairs, (L,R)
For I = 1 To CutNum
 For J = 1 To CutNum
  If I < J Then
   FrNum1 = FrNum1 + 1
   FrL1(FrNum1) = CutPos(I): FrR1(FrNum1) = CutPos(J)
  End If
 Next J
Next I
'***calculate mass for fragments
For I = 1 To FrNum1
 Temp = NINETEEN
 For J = FrL1(I) + 1 To FrR1(I)
  Temp = Temp + AmsAA(J)
 Next J
 If Temp > MsLowLimit And Temp < MsHighLimit Then
  FrNum2 = FrNum2 + 1
  FrMs(FrNum2) = Temp
  FrL2(FrNum2) = FrL1(I): FrR2(FrNum2) = FrR1(I): pNum(FrNum2) = 0
 End If
 '***add pFragments and calculate their mass
 If pSer$ = "S" Or pThr$ = "T" Or pTyr$ = "Y" Then
  pNumTemp = 0
  For K = FrL1(I) + 1 To FrR1(I)
   If AA$(K) = pSer$ Or AA$(K) = pThr$ Or AA$(K) = pTyr$ Then
    pNumTemp = pNumTemp + 1:  Temp2 = Temp + pNumTemp * EIGHTY
    If Temp2 > MsLowLimit And Temp2 < MsHighLimit Then
     pFrNum = pFrNum + 1: FrNum2 = FrNum2 + 1
     FrMs(FrNum2) = Temp2
     FrL2(FrNum2) = FrL1(I): FrR2(FrNum2) = FrR1(I): pNum(FrNum2) = pNumTemp
    End If
   End If
  Next K
 End If
Next I
Form1.Print " total # of unphosphorylated fragments from this protein=", FrNum1
Form1.Print " cal'd # of unphosphorylated fragments within this mass range= ", FrNum2 - pFrNum
Form1.Print " cal'd # of phosphorylated fragments within this mass range= ", pFrNum
Form1.Print " cal'd # of fragments within this mass range= ", FrNum2

'***sort fragments by mass
```

```
  For I = 1 To FrNum2
    MsIdx(I) = I
  Next I
  For I = 2 To FrNum2
    For J = I To 2 Step -1
      If FrMs(MsIdx(J)) < FrMs(MsIdx(J - 1)) Then
        Temp = MsIdx(J): MsIdx(J) = MsIdx(J - 1): MsIdx(J - 1) = Temp
      End If
    Next J
  Next I
  Label12.Visible = 0
End Sub

'*** pull down menu: change current directory browser
Sub CD_Click ()
  Beep
  Drive2.Enabled = 1: Dir2.Enabled = 1: CommandOK.Enabled = 1: Drive2.Visible = 1: Dir2.Visible = 1:
CommandOK.Visible = 1
End Sub

'*** CLEAR buttom
Sub Command1_Click ()
  Beep
  CAANum = 0: Label3.Caption = "": Label4.Caption = ""
End Sub

'*** change current directory browser: OK buttom
Sub CommandOK_Click ()
  Beep
  Drive2.Enabled = 0: Dir2.Enabled = 0: CommandOK.Enabled = 0: Drive2.Visible = 0: Dir2.Visible = 0:
CommandOK.Visible = 0
End Sub

'*** input amino acid resideus which should be cut
Sub CutAA_KeyPress (Keyascii As Integer)
  Beep
  '***accept the rules to cut peptide chain from user
  CAANum = CAANum + 1
  CutAA_$(CAANum) = UCase$(Chr$(Keyascii))
  Temp1 = CutAA_$(CAANum) = "A" Or CutAA_$(CAANum) = "C" Or CutAA_$(CAANum) = "D" Or
CutAA_$(CAANum) = "E" Or CutAA_$(CAANum) = "F" Or CutAA_$(CAANum) = "G" Or
CutAA_$(CAANum) = "H" Or CutAA_$(CAANum) = "I" Or CutAA_$(CAANum) = "K"
  Temp2 = CutAA_$(CAANum) = "L" Or CutAA_$(CAANum) = "M" Or CutAA_$(CAANum) = "N" Or
CutAA_$(CAANum) = "O" Or CutAA_$(CAANum) = "P" Or CutAA_$(CAANum) = "Q" Or
CutAA_$(CAANum) = "R" Or CutAA_$(CAANum) = "S" Or CutAA_$(CAANum) = "T"
  Temp3 = CutAA_$(CAANum) = "V" Or CutAA_$(CAANum) = "W" Or CutAA_$(CAANum) = "Y"
  If Temp1 Or Temp2 Or Temp3 Then
    CutAA.Text = ""
    CutCN$(CAANum) = CutCorN$
    If CutCN$(CAANum) = "Before" Then
      Label3.Caption = Label3.Caption + CutAA_$(CAANum)
    Else
      Label4.Caption = Label4.Caption + CutAA_$(CAANum)
    End If
```

```
    If Chr$(Keyascii) = "/" Then
      CAANum = 0: Label3.Caption = "": Label4.Caption = ""
    End If
  Else
    CAANum = CAANum - 1
  End If
End Sub

'*** menu: Specific Dedradation
Sub Degradation_Click ()
  'inactivate file dir box
  Drive1.Enabled = 0: Dir1.Enabled = 0: File1.Enabled = 0: : Drive1.Visible = 0: Dir1.Visible = 0:
File1.Visible = 0
  Drive2.Enabled = 0: Dir2.Enabled = 0: CommandOK.Enabled = 0: Drive2.Visible = 0: Dir2.Visible = 0:
CommandOK.Visible = 0
End Sub

'*** file browser: change directory
Sub Dir1_Change ()
  File1.Path = Dir1.Path
  ChDir Dir1.Path
End Sub

'*** change current directory browser: change directory
Sub Dir2_Change ()
  ChDir Dir2.Path
End Sub

'*** file browser: change drive
Sub Drive1_Change ()
  Dir1.Path = Drive1.Drive
  ChDrive Drive1.Drive
End Sub

'*** change current directory browser: change drive
Sub Drive2_Change ()
  Dir2.Path = Drive2.Drive
  ChDrive Drive2.Drive
End Sub

'*** menu: File
Sub File_Click ()
  'inactivate file dir box
  Drive1.Enabled = 0: Dir1.Enabled = 0: File1.Enabled = 0: : Drive1.Visible = 0: Dir1.Visible = 0:
File1.Visible = 0
  Drive2.Enabled = 0: Dir2.Enabled = 0: CommandOK.Enabled = 0: Drive2.Visible = 0: Dir2.Visible = 0:
CommandOK.Visible = 0
End Sub

'*** file browser: select file
Sub File1_Click ()
  Beep
  Dim start As Integer
  NumberAA = 0
```

```
On Error GoTo ErrHandler2
'***open file and read amino acid sequence
FileName$ = File1.FileName
Open FileName$ For Input As #1
Line Input #1, peptide_name$
peptide$ = "xxx"
Do Until peptide$ = ""
  Line Input #1, peptide$
  NumberAA = NumberAA + Len(peptide$)
  start = NumberAA - Len(peptide$) + 1
  For I = start To NumberAA
    AA$(I) = Mid$(peptide$, I - start + 1, 1)
  Next I
Loop
Close #1
'***assign average mass to amino acid residues
For I = 1 To NumberAA
  Select Case AA$(I)
    Case "A"
      AmsAA(I) = AMsA
    Case "R"
      AmsAA(I) = AMsR
    Case "N"
      AmsAA(I) = AMsN
    Case "D"
      AmsAA(I) = AMsD
    Case "C"
      AmsAA(I) = AMsC
    Case "E"
      AmsAA(I) = AMsE
    Case "Q"
      AmsAA(I) = AMsQ
    Case "G"
      AmsAA(I) = AMsG
    Case "H"
      AmsAA(I) = AMsH
    Case "I"
      AmsAA(I) = AMsI
    Case "L"
      AmsAA(I) = AMsL
    Case "K"
      AmsAA(I) = AMsK
    Case "M"
      AmsAA(I) = AMsM
    Case "O"
      AmsAA(I) = AMsO
    Case "F"
      AmsAA(I) = AMsF
    Case "P"
      AmsAA(I) = AMsP
    Case "S"
      AmsAA(I) = AMsS
    Case "T"
      AmsAA(I) = AMsT
```

```
    Case "W"
      AmsAA(I) = AMsW
    Case "Y"
      AmsAA(I) = AMsY
    Case "V"
      AmsAA(I) = AMsV
    Case Else
      Form1.Cls: For J = 1 To 17: Form1.Print : Next J
      Form1.Print " Amino Acid Residue #"; I; ", '"; AA$(I); "', cannot be identified, or it's not a sequence
file format.": GoTo LabelAA
    End Select
  Next I
  AmsAAfirst = AmsAA(1): AmsAAlast = AmsAA(NumberAA)
  '***print out amino acid seqence on screen
  Form1.Cls: For I = 1 To 15: Form1.Print : Next I
  Form1.Print " Protein/Peptide : "; peptide_name$
  Form1.Print " Number of Amino Acid= "; NumberAA
  PepMW = NINETEEN
  For I = 1 To NumberAA
    PepMW = PepMW + AmsAA(I)
  Next I
  PepMW = PepMW
  Form1.Print " Peptide Molecular Weight (MH+) = "; PepMW
  Form1.Print " (without terminal modification)"
  Form1.Print
  Form1.Print " 1 ";
  For I = 1 To NumberAA
    Form1.Print AA$(I);
    If (I - Int(I / 50) * 50) = 0 Then
      Form1.Print
      Form1.Print I + 1;
    ElseIf (I - Int(I / 10) * 10) = 0 Then
      Form1.Print " ";
    End If
  Next I
  'activate menu bar
  Macfile.Enabled = 1: PrintSeq.Enabled = 1: CD.Enabled = 1
  CalFrag.Enabled = 1: PrinterCalFrag.Enabled = 0: FileCalFrag.Enabled = 0: ScreenCalFrag.Enabled = 0
  SrchMass.Enabled = 1: SrchSeq.Enabled = 1
  Exit Sub
ErrHandler2:
    Close #1
    Form1.Cls: For I = 1 To 15: Form1.Print : Next I
    Form1.Print " It is not a sequence file format."
    Form1.Print " Error"; Err; ", "; Error$(Err); "."
    Resume LabelAA
LabelAA:
  'deactivate menu
  Macfile.Enabled = 0: PrintSeq.Enabled = 0
  CalFrag.Enabled = 0: PrinterCalFrag.Enabled = 0: FileCalFrag.Enabled = 0: ScreenCalFrag.Enabled = 0
  SrchMass.Enabled = 0: SrchSeq.Enabled = 0
End Sub

'*** pull down menu: print calculation results to a text file
```

```
Sub FileCalFrag_Click ()
 Beep
 Dim tx1$, tx2$, tx3$, tx4$
 '***output the result of CalFrag to a text file
 tx1$ = Dir1.Path
 tx2$ = ""
 For I = 1 To 9
  tx4$ = Mid$(FileName$, I, 1)
  If tx4$ = "." Then
   Exit For
  Else
   tx2$ = tx2$ + tx4$
  End If
 Next I
 tx3$ = tx1$ + "\" + tx2$ + ".frg"
 FileName2$ = InputBox$("FileName to Save? You may save results as a file, so you can manupulate
them. (eg. using a spread sheet, Microsoft Excel, etc..) Comma (code 44) is used as delineator here.", "",
tx3$)
 If FileName2$ = "" Then
  GoTo LabelDD
 End If
 Open FileName2$ For Output As #2
 Print #2, "Peptide Name="; ","; peptide_name$
 Print #2, "Number of Amino Acid = "; ","; NumberAA
 Print #2, " Peptide Molecular Weight (MH+) = "; PepMW
 Print #2, "N-terminal modification="; ","; Nmodify
 Print #2, "C-terminal modification="; ","; Cmodify
 Print #2, "Cleave at:"
 TempStr$ = ""
 For I = 1 To CAANum
  If CutCN$(I) = "Before" Then
   TempStr$ = TempStr$ + CutAA_$(I)
  End If
 Next I
 Print #2, "N-terminal of :"; ","; TempStr$
 TempStr$ = ""
 For I = 1 To CAANum
  If CutCN$(I) = "After" Then
   TempStr$ = TempStr$ + CutAA_$(I)
  End If
 Next I
 Print #2, "N-terminal of "; ","; TempStr$
 TempStr$ = pSer$ + pThr$ + pTyr$
 If TempStr$ = "" Then
  TempStr$ = "none"
 End If
 Print #2, "Phosphorylation at:"; ","; TempStr$
 TempStr$ = Str$(MsLowLimit) + " - " + Str$(MsHighLimit)
 Print #2, "mass range considered:"; ","; TempStr$
 Print #2, "total # of unphosphorylated fragments from this protein="; ","; FrNum1
 Print #2, "cal'd # of unphosphorylated fragments within this mass range="; ","; FrNum2 - pFrNum
 Print #2, "cal'd # of phosphorylated fragments within this mass range="; ","; pFrNum
 Print #2, "cal'd # of fragments within this mass range="; ","; FrNum2
 Print #2, "mass"; ","; "# of phosphorylation"; ","; "start residue #"; ","; "end residue #"
```

```
  For I = 1 To FrNum2
    Print #2, Format(FrMs(MsIdx(I)), "0.00"); ","; pNum(MsIdx(I)); ","; FrL2(MsIdx(I)) + 1; ",";
FrR2(MsIdx(I))
  Next I
  Close #2
LabelDD:
End Sub

'*** main form for this program
Sub Form_Load ()
  WindowState = 2
  '***assign initial values to some variables
  Nmodify = 0: Cmodify = 0
  CutCorN$ = "After"
  pSer$ = "": pThr$ = "": pTyr$ = ""
  SeqnSrch$ = "R?S"
  '***disable items in menu
  Drive1.Enabled = 0: Dir1.Enabled = 0: File1.Enabled = 0: Drive1.Visible = 0: Dir1 Visible = 0:
File1.Visible = 0
  Drive2.Enabled = 0: Dir2.Enabled = 0: CommandOK.Enabled = 0: Drive2.Visible = 0: Dir2.Visible = 0:
CommandOK.Visible = 0
  Macfile.Enabled = 0: PrintSeq.Enabled = 0
  CalFrag.Enabled = 0: PrinterCalFrag.Enabled = 0: FileCalFrag.Enabled = 0: ScreenCalFrag.Enabled = 0
  SrchMass.Enabled = 0: SrchSeq.Enabled = 0
  Label12.Visible = 0
End Sub

'*** menu: Help
Sub Help_Click ()
  'inactivate file dir box
  Drive1.Enabled = 0: Dir1.Enabled = 0: File1.Enabled = 0: : Drive1.Visible = 0: Dir1.Visible = 0:
File1.Visible = 0
  Drive2.Enabled = 0: Dir2.Enabled = 0: CommandOK.Enabled = 0: Drive2.Visible = 0: Dir2.Visible = 0:
CommandOK.Visible = 0
End Sub

'*** pull down menu of Help: How to build a protein sequence file
Sub How_Click ()
  Beep: Form1.Cls: For I = 1 To 12: Form1.Print : Next I
  Form1.Print " The file which contains protein sequence has to"
  Form1.Print " be the format as follows, so it can be opened by"
  Form1.Print " this program. A demo file with filename bcasein.txt"
  Form1.Print " should be sent to you with this program."
  Form1.Print
  Form1.Print "   (name of protein)[CR,L]"
  Form1.Print "   (any # of amino acid sequence)[CR,L]"
  Form1.Print "   (any # of amino acid sequence)[CR,L]"
  Form1.Print "      ....."
  Form1.Print "      ....."
  Form1.Print "   (any # of amino acid sequence)[CR,L]"
  Form1.Print "   [CR,L]"
  Form1.Print "   (anything beyond this point will not be read)"
  Form1.Print
  Form1.Print " [CR,L]: carriage return (code 13) and line feed"
```

```
  Form1.Print " (code 10), two successive [CR,L]'s denote the end"
  Form1.Print " of amino acid sequence."
  Form1.Print
  Form1.Print " For example: the sequence of angiotensin I from"
  Form1.Print " salmon is NRVYVHPFNL"
  Form1.Print
  Form1.Print "   angiotensin I, salmon[CR,L]"
  Form1.Print "   NRVYVH[CR,L]"
  Form1.Print "   PFNL[CR,L]"
  Form1.Print "   [CR,L]"
  Form1.Print
  Form1.Print " You may use a word processor(e.g. Microsoft Word)"
  Form1.Print " or text editor to build the protein sequence file,"
  Form1.Print " or trim an existing file which containing protein"
  Form1.Print " sequence from other source."
  MsgBox "", 0, ""
  Form1.Cls
End Sub

'*** pull down menu of Help: How to export a MacProMass file format
Sub How2_Click ()
  Beep: Form1.Cls: For I = 1 To 12: Form1.Print : Next I
  Form1.Print " You may like to export this protein sequence"
  Form1.Print " to a Macintosh file, so MacProMass can read it."
  Form1.Print " Follow the steps:"
  Form1.Print
  Form1.Print " 1. From 'File' menu, use 'Convert to MacProMass"
  Form1.Print "  Format'. A DOS file with extension '.M~' will"
  Form1.Print "  be generated."
  Form1.Print
  Form1.Print " 2. Use 'MAC-IN-DOS' software to convert DOS file"
  Form1.Print "  format to Macintoch file format. Execute 'CLINK.EXE'."
  Form1.Print "  Follow the instructions. Use binary copy."
  Form1.Print " 'MAC-IN-DOS' will allow you convert DOS file to"
  Form1.Print "  Mac-formatted disk using a DOS floppy drive."
  Form1.Print "  'MAC-IN-DOS' can only use 1.44 MB formatted disk."
  Form1.Print "  'MAC-IN-DOS' is a product of Pacific Microelec-"
  Form1.Print "  tronics, Inc. Mountain View, California,"
  Form1.Print "  phone number:(415)9486200."
  MsgBox "", 0, ""
  Form1.Cls
End Sub

'*** label
Sub Label1_Click ()

End Sub

'*** label
Sub Label10_Click ()

End Sub

'*** label
```

```
Sub Label11_Click ()

End Sub

'*** label
Sub Label12_Click ()

End Sub

'*** label
Sub Label13_Click ()

End Sub

'*** label
Sub Label14_Click ()

End Sub

'*** label
Sub Label15_Click ()

End Sub

'*** label
Sub Label2_Click ()

End Sub

'*** label
Sub Label3_Click ()

End Sub

'*** label
Sub Label4_Click ()

End Sub

'*** label
Sub Label5_Click ()

End Sub

'*** label
Sub Label6_Click ()

End Sub

'*** label
Sub Label7_Click ()

End Sub
```

```
'*** label
Sub Label8_Click ()

End Sub

'*** label
Sub Label9_Click ()

End Sub

'*** pull down menu: convert sequence file to MacProMass format using Comlink
Sub Macfile_Click ()
  Beep
  tx1$ = Dir1.Path
  tx2$ = "": tx4$ = ""
  For I = 1 To 9
    tx4$ = Mid$(FileName$, I, 1)
    If tx4$ = "." Then
      Exit For
    End If
    tx2$ = tx2$ + tx4$
  Next I
  tx3$ = tx1$ + "\" + tx2$ + ".M~"
  FileName3$ = InputBox$("FileName to Save? must use file extension: '.M~'", "", tx3$)
  If FileName3$ = "" Then
    GoTo LabelCC
  End If
  CR$ = Chr$(13): EF$ = Chr$(10): tx1$ = ",0,0,0,0,0,0,0,0,0,0,0,0"
  tx2$ = "Hydrogen,1.0078,1.0079,0,1,0,0,0,0,0,0,0,0"
  tx3$ = "Free Acid,17.0072,17.0073,0,1,0,1,0,0,0,0,0,0"
  Open FileName3$ For Output As #3
  For I = 1 To NumberAA
    Print #3, AA$(I);
  Next I
  Print #3, CR$;
  For I = 1 To 6
    Print #3, tx1$; CR$;
  Next I
  Print #3, tx2$; CR$; tx3$; CR$; EF$
  Close #3
LabelCC:
End Sub

'*** inputbox: specify mass range, high
Sub MassHigh_Change ()
  Beep
  '*** accept high mass limit from user
  MsHighLimit = Val(MassHigh.Text)
End Sub

'*** inputbox: specify mass range, low
Sub MassLow_Change ()
  Beep
  '*** accept high mass limit from user
```

```
  MsLowLimit = Val(MassLow.Text)
End Sub

'*** inputbox: define N-terminal modification by mass
Sub N_modify_Change ()
  Beep
  Nmodify = Val(N_modify.Text)
End Sub

'*** pull down menu of Help: a note from programmer
Sub Note_Click ()
  Beep: Form1.Cls: For I = 1 To 12: Form1.Print : Next I
  Form1.Print " I have tried to make this program easy to use."
  Form1.Print " However, I am not a professional programmer."
  Form1.Print " If you find any difficulty to use it, or any"
  Form1.Print " bugs which cause troubles, please contact me"
  Form1.Print " by phone or mail. I will be more than"
  Form1.Print " happy to try to help or solve troubles."
  Form1.Print
  Form1.Print "   Pao-Chi Liao"
  Form1.Print "   (517) 353-0612"
  Form1.Print "   Mass Spectrometry Facility, Biochemistry "
  Form1.Print "   Michigan State University"
  Form1.Print "   East Lansing, Michigan 48824"
  MsgBox "", 0, ""
  Form1.Cls
End Sub

'*** pull down menu: open sequence file browser
Sub OpenFile_Click ()
  Beep
  'activate file bar
  Drive1.Enabled = 1: Dir1.Enabled = 1: File1.Enabled = 1: Drive1.Visible = 1: Dir1.Visible = 1:
File1.Visible = 1
  'inactivate menu
  Macfile.Enabled = 0: PrintSeq.Enabled = 0
  CalFrag.Enabled = 0: PrinterCalFrag.Enabled = 0: FileCalFrag.Enabled = 0: ScreenCalFrag.Enabled = 0
  SrchMass.Enabled = 0: SrchSeq.Enabled = 0
  'temptrial
  ValSrchMs = 2000: RngSrchMs = 2: MsLowLimit = Val(MassLow.Text): MsHighLimit =
Val(MassHigh.Text)
End Sub

'*** pull down menu: print calculation results to printer
Sub PrinterCalFrag_Click ()
  Beep
  '***print out amino acid seqence and termini modifications to printer
  Printer.Print "Protein/Peptide : "; peptide_name$
  Printer.Print " 1 ";
  For I = 1 To NumberAA
    Printer.Print AA$(I);
    If (I - Int(I / 50) * 50) = 0 Then
      Printer.Print
      Printer.Print I + 1;
```

```
  ElseIf (I - Int(I / 10) * 10) = 0 Then
   Printer.Print " ";
  End If
 Next I
 Printer.Print
 Printer.Print "total number of amino acid = "; NumberAA
 Printer.Print "peptide molecular weight (MH+) = "; PepMW
 Printer.Print "N-terminal modification = "; Nmodify
 Printer.Print "C-terminal modification = "; Cmodify
 '***print out cut positions and phosphorylation sites considered to printer
 Printer.Print "Cleave at: "; Tab(12); "N-terminal of ";
 For I = 1 To CAANum
  If CutCN$(I) = "Before" Then
   Printer.Print CutAA_$(I);
  End If
 Next I
 Printer.Print
 Printer.Print Tab(12); "C-terminal of ";
 For I = 1 To CAANum
  If CutCN$(I) = "After" Then
   Printer.Print CutAA_$(I);
  End If
 Next I
 Printer.Print
 Printer.Print "Phosphorylation at: "; pSer$; pThr$; pTyr$;
 If pSer$ = "" And pThr$ = "" And pTyr$ = "" Then
  Printer.Print "none"
 Else
  Printer.Print
 End If
 '***print out mass range considered to printer
 Printer.Print "mass range considered: "; MsLowLimit; " - "; MsHighLimit
 '***print out fragments to printer
 Printer.Print
 Printer.Print "total # of unphosphorylated fragments from this protein=", FrNum1
 Printer.Print "cal'd # of unphosphorylated fragments within this mass range= ", FrNum2 - pFrNum
 Printer.Print "cal'd # of phosphorylated fragments within this mass range= ", pFrNum
 Printer.Print "cal'd # of fragments within this mass range= ", FrNum2
 Printer.Print
 Printer.Print Tab(3); "[M+H]+, m/z"; Tab(24); "peptide"
 Printer.Print Tab(1); "----------"; Tab(18); "----------"
 For I = 1 To FrNum2
  Printer.Print Tab(3); Format(FrMs(MsIdx(I)), "0.00"); Tab(20), pNum(MsIdx(I)); "p,"; FrL2(MsIdx(I))
+ 1; " - "; FrR2(MsIdx(I))
  If (I - Int(I / 10) * 10) = 0 Then
   Printer.Print " "
  End If
 Next I
 Printer.Print Tab(1); "----------"; Tab(18); "----------"
 Printer.EndDoc
End Sub

'*** pull down menu: print protein sequence and molecular weight
Sub PrintSeq_Click ()
```

```
  Beep
  '***print out amino acid seqence and termini modifications to printer
  Printer.Print "Protein/Peptide : "; peptide_name$
  Printer.Print " 1 ";
  For I = 1 To NumberAA
    Printer.Print AA$(I);
    If (I - Int(I / 50) * 50) = 0 Then
      Printer.Print
      Printer.Print I + 1;
    ElseIf (I - Int(I / 10) * 10) = 0 Then
      Printer.Print " ";
    End If
  Next I
  Printer.Print
  Printer.Print "total number of amino acid = "; NumberAA
  Printer.Print "peptide molecular weight (MH+) = "; PepMW
  Printer.Print "N-terminal modification = "; Nmodify
  Printer.Print "C-terminal modification = "; Cmodify
  Printer.EndDoc
End Sub

'*** pull down menu: print calculation results on screen
Sub ScreenCalFrag_Click ()
  Beep
  '***print sorted (by mass) Fragments on screen
  Form1.Print
  Form1.Cls
  For I = 1 To FrNum2
    Form1.Print " [M+H]+,m/z="; Format(FrMs(MsIdx(I)), "0.00"); Tab(24); " --- "; Tab(29);
pNum(MsIdx(I)); "p,"; FrL2(MsIdx(I)) + 1; " - "; FrR2(MsIdx(I))
    If (I - Int(I / 40) * 40) = 0 Then
      MsgBox ("Continue!")
      Form1.Cls
    ElseIf (I - Int(I / 10) * 10) = 0 Then
      Form1.Print " "
    End If
  Next I
End Sub

'*** menu: Search
Sub Search_Click ()
  'inactivate file dir box
  Drive1.Enabled = 0: Dir1.Enabled = 0: File1.Enabled = 0: : Drive1.Visible = 0: Dir1.Visible = 0:
File1.Visible = 0
  Drive2.Enabled = 0: Dir2.Enabled = 0: CommandOK.Enabled = 0: Drive2.Visible = 0: Dir2.Visible = 0:
CommandOK.Visible = 0
End Sub

'*** check box: if phosphorylation on serine is considered
Sub Ser_Click ()
  Beep
  '***accept checkbox from user if considering phosphorylation on serine
  If Ser.Value Then
    pSer$ = "S"
```

```
  Else
    pSer$ = ""
  End If
End Sub

'*** pull down menu: search possible peptide sequence for certain mass range, regardless specific
cleavage
Sub SrchMass_Click ()
  Beep
  CountL = 1: CountR = Int(ValSrchMs / 200): SrchMsTemp = NINETEEN: CountMtch = 0
  For I = 1 To CountR
    SrchMsTemp = SrchMsTemp + AmsAA(I)
  Next I
  Do While CountR < NumberAA
    Do While (ValSrchMs - SrchMsTemp) > RngSrchMs
      CountR = CountR + 1
      If CountR = NumberAA Then
        SrchMsTemp = SrchMsTemp + AmsAA(CountR)
        Exit Do
      End If
      SrchMsTemp = SrchMsTemp + AmsAA(CountR)
    Loop
    If CountR = NumberAA Then
      Exit Do
    End If
    Do While Abs(SrchMsTemp - ValSrchMs) < RngSrchMs
      CountMtch = CountMtch + 1: MsSrchMtch(CountMtch) = SrchMsTemp
      LSrchMtch(CountMtch) = CountL: RSrchMtch(CountMtch) = CountR
      CountR = CountR + 1: SrchMsTemp = SrchMsTemp + AmsAA(CountR)
    Loop
    SrchMsTemp = SrchMsTemp - AmsAA(CountL): CountL = CountL + 1
    Do While (ValSrchMs - SrchMsTemp) < RngSrchMs
      SrchMsTemp = SrchMsTemp - AmsAA(CountR): CountR = CountR - 1
    Loop
  Loop
  Do While ValSrchMs - SrchMsTemp < RngSrchMs
    If Abs(ValSrchMs - SrchMsTemp) < RngSrchMs Then
      CountMtch = CountMtch + 1: MsSrchMtch(CountMtch) = SrchMsTemp
      LSrchMtch(CountMtch) = CountL: RSrchMtch(CountMtch) = CountR
    End If
    SrchMsTemp = SrchMsTemp - AmsAA(CountL): CountL = CountL + 1
    Form1.Print "******"
  Loop
  'print search results on screen
  Form1.Cls
  For I = 1 To CountMtch
    Form1.Print "m/z= "; Tab(10); Format(MsSrchMtch(I), "0.00"); Tab(20); " -- "; Tab(24); "(";
AA$(LSrchMtch(I) - 1); ")"; Str$(LSrchMtch(I)); AA$(LSrchMtch(I)); " -"; Str$(RSrchMtch(I));
AA$(RSrchMtch(I)); " ("; AA$(RSrchMtch(I) + 1); ")"
    If (I - Int(I / 40) * 40) = 0 Then
      MsgBox ("Continue!")
      Form1.Cls
    ElseIf (I - Int(I / 10) * 10) = 0 Then
      Form1.Print " "
```

```
  End If
 Next I
 If CountMtch = 0 Then
  Form1.Print "There is no match."
 Else
 'print to printer
 TempStr$ = InputBox$("Print to printer?", , "Yes.")
 If TempStr$ = "Yes." Then
  Printer.Print "Protein/Peptide : "; peptide_name$
  Printer.Print " 1 ";
  For I = 1 To NumberAA
   Printer.Print AA$(I);
   If (I - Int(I / 50) * 50) = 0 Then
    Printer.Print
    Printer.Print I + 1;
   ElseIf (I - Int(I / 10) * 10) = 0 Then
    Printer.Print " ";
   End If
  Next I
  Printer.Print
  Printer.Print "total number of amino acid = "; NumberAA
  Printer.Print "peptide molecular weight (MH+) = "; PepMW
  Printer.Print "N-terminal modification = "; Nmodify
  Printer.Print "C-terminal modification = "; Cmodify
  Printer.Print
  Printer.Print "Peptide sequences with m/z of [M+H]+ between "; Format(ValSrchMs - RngSrchMs,
"0.00"); " and "; Format(ValSrchMs + RngSrchMs, "0.00"); " are:"
  For I = 1 To CountMtch
   Printer.Print "m/z= "; Tab(11); Format(MsSrchMtch(I), "0.00"); Tab(21); " -- "; Tab(25); "(";
AA$(LSrchMtch(I) - 1); ")"; Str$(LSrchMtch(I)); AA$(LSrchMtch(I)); " -"; Str$(RSrchMtch(I));
AA$(RSrchMtch(I)); " ("; AA$(RSrchMtch(I) + 1); ")"
  Next I
  Printer.EndDoc
 End If
End If
End Sub

'*** inputbox: specify mass range for search
Sub SrchMsRng_Change ()
 Beep
 '*** accept mass range for search
 RngSrchMs = Val(SrchMsRng.Text)
End Sub

'*** inputbox: specify mass value for search
Sub SrchMsVal_Change ()
 Beep
 '*** accept mass value for search
 ValSrchMs = Val(SrchMsVal.Text)
End Sub

'*** pull down menu: search for certain peptide sequence("?", for any amino acid), regardless specific
cleavage
Sub SrchSeq_Click ()
```

```
Beep
Dim Lseq As Integer, Temp2 As Integer
ReDim Seq$(20)
Dim There_is_match As Integer
There_is_match = 0
 Lseq = Len(SeqnSrch$)
For I = 1 To Lseq
  Seq$(I) = Mid$(SeqnSrch$, I, 1)
Next I
Form1.Cls
For I = 1 To NumberAA - Lseq + 1
  Temp2 = 0
  For J = 1 To Lseq
    If AA$(I + J - 1) = Seq$(J) Or Seq$(J) = "?" Then
      Temp2 = Temp2 + 1
    End If
  Next J
  If Temp2 = Lseq Then
    There_is_match = 1
    Form1.Print Tab(3); I; " -"; I + Lseq - 1; ", "; Tab(17); "--"; Tab(20);
    For K = I To I + Lseq - 1: Form1.Print AA$(K); : Next K
    Form1.Print
  End If
Next I
If There_is_match = 0 Then
  Form1.Print "There is no match."
Else
 'print to printer
 TempStr$ = InputBox$("Print to printer?", , "Yes.")
 If TempStr$ = "Yes." Then
   Printer.Print "Protein/Peptide : "; peptide_name$
   Printer.Print " 1 ";
   For I = 1 To NumberAA
     Printer.Print AA$(I);
     If (I - Int(I / 50) * 50) = 0 Then
       Printer.Print
       Printer.Print I + 1;
     ElseIf (I - Int(I / 10) * 10) = 0 Then
       Printer.Print " ";
     End If
   Next I
   Printer.Print
   Printer.Print "total number of amino acid = "; NumberAA
   Printer.Print "peptide molecular weight (MH+) = "; PepMW
   Printer.Print "N-terminal modification = "; Nmodify
   Printer.Print "C-terminal modification = "; Cmodify
   Printer.Print
   Printer.Print "Peptides with consensus sequence, "; SeqnSrch$; ", are:"
   For I = 1 To NumberAA - Lseq + 1
     Temp2 = 0
     For J = 1 To Lseq
       If AA$(I + J - 1) = Seq$(J) Or Seq$(J) = "?" Then
         Temp2 = Temp2 + 1
       End If
```

```
    Next J
    If Temp2 = Lseq Then
      Printer.Print Tab(3); I; " -"; I + Lseq - 1; ", "; Tab(17); "--"; Tab(20);
      For K = I To I + Lseq - 1: Printer.Print AA$(K); : Next K
      Printer.Print
    End If
   Next I
   Printer.EndDoc
  End If
 End If
End Sub

'*** inputbox: specify amino acid sequence for search, "?", for any amino acid
Sub SrchSeqn_Change ()
 Beep
 '*** accept mass sequence for search
 SeqnSrch$ = SrchSeqn.Text
End Sub

'*** check box: if phosphorylation on threonine is considered
Sub Thr_Click ()
 Beep
 '***accept checkbox from user if considering phosphorylation on threonine
 If Thr.Value Then
  pThr$ = "T"
 Else
  pThr$ = ""
 End If
End Sub

'*** check box: if phosphorylation on tyrosine is considered
Sub Tyr_Click ()
 Beep
 '***accept checkbox from user if considering phosphorylation on tyrosine
 If Tyr.Value Then
  pTyr$ = "Y"
 Else
  pTyr$ = ""
 End If
End Sub

'*** layouts of form , menu, and objects
Begin Form Form1
  BackColor      = &H00FFFFFF&
  Caption        = "MSU MassMap v.1.1"
  ClientHeight   = 7140
  ClientLeft     = 105
  ClientTop      = 1710
  ClientWidth    = 11700
  Height         = 7830
  Icon           = MSUV11D.FRX:0000
  Left           = 45
  LinkMode       = 1 'Source
  LinkTopic      = "Form1"
```

```
ScaleHeight    = 7140
ScaleWidth     = 11700
Top            = 1080
Width          = 11820
Begin TextBox SrchSeqn
  Height       = 360
  Left         = 7800
  TabIndex     = 28
  Text         = "R?S"
  Top          = 2535
  Width        = 3540
End
Begin TextBox SrchMsRng
  Height       = 360
  Left         = 8460
  TabIndex     = 26
  Text         = "2"
  Top          = 2160
  Width        = 1110
End
Begin TextBox SrchMsVal
  Height       = 360
  Left         = 6360
  TabIndex     = 24
  Text         = "2000"
  Top          = 2160
  Width        = 1140
End
Begin CheckBox Tyr
  Caption      = "Tyrosine"
  Height       = 225
  Left         = 9240
  TabIndex     = 7
  Top          = 1530
  Width        = 1185
End
Begin CheckBox Thr
  Caption      = "Threonine"
  Height       = 210
  Left         = 8040
  TabIndex     = 8
  Top          = 1530
  Width        = 1200
End
Begin CheckBox Ser
  Caption      = "Serine"
  Height       = 255
  Left         = 7080
  TabIndex     = 6
  Top          = 1500
  Width        = 855
End
Begin TextBox C_modify
  Height       = 360
```

```
  Left        = 9240
  TabIndex    = 18
  Text        = "0"
  Top         = 1170
  Width       = 510
End
Begin TextBox N_modify
  Height      = 360
  Left        = 7560
  TabIndex    = 17
  Text        = "0"
  Top         = 1200
  Width       = 495
End
Begin TextBox MassHigh
  Height      = 360
  Left        = 8160
  TabIndex    = 11
  Text        = "3000"
  Top         = 810
  Width       = 900
End
Begin TextBox MassLow
  Height      = 360
  Left        = 6780
  TabIndex    = 10
  Text        = "0"
  Top         = 810
  Width       = 885
End
Begin OptionButton After
  Caption     = "Cleave at C-terminus of:"
  Height      = 255
  Left        = 6240
  TabIndex    = 2
  Top         = 480
  Value       = -1 'True
  Width       = 2340
End
Begin CommandButton Command1
  Caption     = "CLEAR"
  Height      = 255
  Left        = 4920
  TabIndex    = 31
  Top         = 510
  Width       = 675
End
Begin DirListBox Dir2
  Height      = 1605
  Left        = 360
  TabIndex    = 30
  Top         = 480
  Width       = 2775
End
```

```
Begin DirListBox Dir1
  Height        = 1605
  Left          = 120
  TabIndex      = 21
  Top           = 480
  Width         = 2535
End
Begin TextBox CutAA
  Height        = 360
  Left          = 5760
  TabIndex      = 0
  Top           = 345
  Width         = 330
End
Begin OptionButton Before
  BackColor     = &H00FFFFFF&
  Caption       = "Cleave at N-terminusof:"
  Height        = 240
  Left          = 6240
  TabIndex      = 1
  Top           = 270
  Width         = 2325
End
Begin CommandButton CommandOK
  Caption       = "OK"
  Height        = 375
  Left          = 2640
  TabIndex      = 32
  Top           = 120
  Width         = 495
End
Begin DriveListBox Drive2
  Height        = 315
  Left          = 360
  TabIndex      = 29
  Top           = 120
  Width         = 2175
End
Begin DriveListBox Drive1
  Height        = 315
  Left          = 120
  TabIndex      = 20
  Top           = 120
  Width         = 2535
End
Begin FileListBox File1
  Height        = 1950
  Left          = 2640
  TabIndex      = 22
  Top           = 120
  Width         = 1575
End
Begin Line Line10
  X1            = 6480
```

```
  X2          = 11400
  Y1          = 2040
  Y2          = 2040
End
Begin Line Line9
  X1          = 11400
  X2          = 11400
  Y1          = 2880
  Y2          = 2040
End
Begin Line Line8
  X1          = 4320
  X2          = 11400
  Y1          = 2880
  Y2          = 2880
End
Begin Line Line7
  X1          = 4320
  X2          = 4320
  Y1          = 2040
  Y2          = 2880
End
Begin Line Line6
  X1          = 4320
  X2          = 4680
  Y1          = 2040
  Y2          = 2040
End
Begin Label Label5
  AutoSize      = -1 'True
  Caption       = "Search Parameters"
  Height        = 195
  Left          = 4800
  TabIndex      = 34
  Top           = 1920
  Width         = 1620
End
Begin Line Line5
  X1          = 7680
  X2          = 11400
  Y1          = 120
  Y2          = 120
End
Begin Line Line4
  X1          = 4320
  X2          = 4680
  Y1          = 120
  Y2          = 120
End
Begin Label Label2
  AutoSize      = -1 'True
  Caption       = "Specific Degradation Parameters"
  Height        = 195
  Left          = 4800
```

```
  TabIndex      = 33
  Top           = 0
  Width         = 2805
End
Begin Line Line3
  X1            = 4320
  X2            = 4320
  Y1            = 120
  Y2            = 1800
End
Begin Line Line2
  BorderColor   = &H00000000&
  X1            = 11400
  X2            = 11400
  Y1            = 120
  Y2            = 1800
End
Begin Line Line1
  X1            = 4320
  X2            = 11400
  Y1            = 1800
  Y2            = 1800
End
Begin Label Label13
  AutoSize      = -1 'True
  Caption       = "Search for : (a) mass ="
  Height        = 195
  Left          = 4380
  TabIndex      = 23
  Top           = 2280
  Width         = 1950
End
Begin Label Label7
  AutoSize      = -1 'True
  Caption       = "Calculate for Mass Range:"
  Height        = 225
  Left          = 4440
  TabIndex      = 12
  Top           = 840
  Width         = 2310
End
Begin Label Label12
  AutoSize      = -1 'True
  BackColor     = &H00C0C000&
  BorderStyle   = 1 'Fixed Single
  Caption       = " Sorting . It may be slow!"
  FontBold      = -1 'True
  FontItalic    = -1 'True
  FontName      = "MS Sans Serif"
  FontSize      = 17.25
  FontStrikethru = 0 'False
  FontUnderline = 0 'False
  ForeColor     = &H000000C0&
  Height        = 465
```

```
  Left        = 6960
  TabIndex    = 19
  Top         = 3000
  Width       = 4530
End
Begin Label Label15
  AutoSize    = -1 'True
  Caption     = "(b) consensus sequence ="
  Height      = 195
  Left        = 5460
  TabIndex    = 27
  Top         = 2640
  Width       = 2250
End
Begin Label Label14
  AutoSize    = -1 'True
  Caption     = ", range ="
  Height      = 195
  Left        = 7620
  TabIndex    = 25
  Top         = 2280
  Width       = 780
End
Begin Label Label6
  AutoSize    = -1 'True
  Caption     = "Possible Phosphorylations at:"
  Height      = 195
  Left        = 4440
  TabIndex    = 9
  Top         = 1530
  Width       = 2505
End
Begin Label Label10
  AutoSize    = -1 'True
  Caption     = "C-terminus:"
  Height      = 195
  Left        = 8220
  TabIndex    = 15
  Top         = 1200
  Width       = 960
End
Begin Label Label11
  AutoSize    = -1 'True
  Caption     = "N-terminus:"
  Height      = 195
  Left        = 6540
  TabIndex    = 16
  Top         = 1200
  Width       = 975
End
Begin Label Label9
  AutoSize    = -1 'True
  Caption     = "Terminal Modifications:"
  Height      = 195
```

```
  Left        = 4440
  TabIndex    = 14
  Top         = 1200
  Width       = 1980
End
Begin Label Label8
  AutoSize    = -1 'True
  Caption     = "To"
  Height      = 195
  Left        = 7800
  TabIndex    = 13
  Top         = 840
  Width       = 240
End
Begin Label Label4
  Height      = 225
  Left        = 8640
  TabIndex    = 5
  Top         = 510
  Width       = 2250
End
Begin Label Label3
  Height      = 210
  Left        = 8640
  TabIndex    = 4
  Top         = 300
  Width       = 2205
End
Begin Label Label1
  AutoSize    = -1 'True
  Caption     = "Enter Residue:"
  Height      = 195
  Left        = 4425
  TabIndex    = 3
  Top         = 315
  Width       = 1275
End
Begin Menu File
  Caption     = "&File"
  Begin Menu CD
    Caption   = "&Change Current Directory"
  End
  Begin Menu OpenFile
    Caption   = "&Open Sequence File"
  End
  Begin Menu PrintSeq
    Caption   = "&Print Protein Sequence"
  End
  Begin Menu Macfile
    Caption   = "Convert to MacProMass Format/via Comlink"
  End
End
Begin Menu Degradation
  Caption     = "Specific &Degradation"
```

```
    Begin Menu CalFrag
      Caption       =  "&Calculate Fragments"
    End
    Begin Menu ScreenCalFrag
      Caption       =  "Display Fragments on &Screen"
    End
    Begin Menu PrinterCalFrag
      Caption       =  "&Print Calculate Result to Printer"
    End
    Begin Menu FileCalFrag
      Caption       =  "Print Calculate Result to a *.txt File"
    End
  End
  Begin Menu Search
    Caption       =  "&Search"
    Begin Menu SrchMass
      Caption       =  "Sequence of certain &Mass"
    End
    Begin Menu SrchSeq
      Caption       =  "&Consensus Sequence"
    End
  End
  Begin Menu Help
    Caption       =  "&Help"
    Begin Menu note
      Caption       =  "A note from programmer"
    End
    Begin Menu about
      Caption       =  "About this program"
    End
    Begin Menu How
      Caption       =  "How to build protein sequence file"
    End
    Begin Menu How2
      Caption       =  "How to export a MacProMass fileformat"
    End
  End
End
```

Appendix VII - Instruction for the use of MSU MassMap v.1.1

June 13, 1994

The text of this instruction is contained in the file 'readme.txt' in the sent floppy disk. You may use Microsoft Word to read it. It is saved in the text format.

Installation of this program:

To install MSU MassMap v.1.1, please run (in Windows) the file 'setup.exe' which is in the sent floppy disk. Follow instructions. Two files, 'msuv11.exe' and 'bcasein.txt' will be generated in the c:\msuv11 directory (or, as you specify.). A MSU MassMap v.1.1 icon will be generated in the Windows. Run this icon. There are some information in HELP to assist you in the use of this program. If the window layout of this program does not fit your monitor screen, go to 'Windows Setup', 'change system setting', 'Display', change the setting to 'Super VGA (800x600, 16 colors, small fonts)'.

What does this program do?

This program was written by Pao-Chi Liao at the Mass Spectrometry Facility, Michigan State University, to calculate the average masses of possible peptide fragments and phosphopeptide fragments from specific enzymatic or chemical degradation of a protein, and the m/z values of the mass spectral peaks for the corresponding [M+H]+ ions. In this program, 'degradation' merely means hydrolysis. The average masses were used instead of monoisotopic masses to make the program suitable for the mass mapping analyses for MALDI-TOF-MS measurements of proteins or peptides. The program computes masses for peptides from not only complete but also partial digestion. For a detailed discussion on how to locate phosphorylation sites in a phosphoprotein using MALDI-TOF-MS and mass mapping, see P.-C. Liao, J. Leykam, P. Andrews, D. Gage and J. Allison, ANALYTICAL BIOCHEMISTRY, 219, 9-20 (1994).

This program allows you define the rules of specific degradation. You may define the specific cleavage sites are either at the N-terminus or the C-terminus of some residues. The program applies the rules defined by you to a protein sequence and generates all the possible peptides from complete and partial digestion. The calculation also includes possible phosphorylation at serine, threonine or tyrosine, if you specify. In the next step, all the corresponding masses of generated peptides and phosphopeptides are calculated, and their corresponding m/z values of $[M+H]^+$ are reported. A list which contains generated peptides sorted according to their masses will be reported. To avoid a lengthy list, you may specify the high and low mass limit. Only peptides with the masses within the specified mass range will be generated. You may also define an N-terminus or C-terminus modification to a protein sequence.

You may also search a peptide sequence of certain mass or consensus sequence within a protein sequence. In this case, no specific degradation is considered. The program searches all possible peptide sequences along the protein sequence. You may also export a protein sequence to a file which can be converted to Macintosh file format, so you can utilize the functions in MacProMass software (by Terry Lee and Sunil Vemuri, Beckman Research Institute of the City of Hope, Durate, CA).

Run this program:

1. First, you need a protein sequence. This program does not have the function to manually input a protein sequence and build a protein sequence file, because this can be easily done by a word processor (e.g. Microsoft Word) or a text editor.

The file which contains the protein sequence has to be the format as follows, so it can be opened by this program. A demo file with filename bcasein.txt should be sent to you with this program.

(name of protein)[CR,L]
(any # of amino acid sequence)[CR,L]
(any # of amino acid sequence)[CR,L]
..... .
.....
(any # of amino acid sequence)[CR,L]
[CR,L]
(anything beyond this point will not be read)"

[CR,L] denotes carriage return (code 13) and line feed (code 10), two successive [CR,L] denote end of amino acid sequence.

For example: the sequence of angiotensin I from salmon is NRVYVHPFNL"

angiotensin I, salmon[CR,L]
NRVYVH[CR,L]
PFNL[CR,L]
[CR,L]

Frequently, you may trim an existing file which contains a protein sequence from other source.

2. Run 'MSU MassMap v.1.1' icon in Windows. Use 'Open' in 'File' menu to open the protein sequence file which you have built or the demo file, 'bcasein.txt', which lists the sequence of bovine beta-casein. Click the filename to open it. Once the protein sequence file is opened, it is displayed on the screen. If you fail to open the protein sequence file which you have built, check the file format again.

3. Specify/change the parameters listed in ' Specific Degradation Parameters' box. You may enter multiple amino acid residues at which either their N-terminus or C-terminus will be specifically hydrolyzed. Use 'CLEAR' buttom to start with new amino acid residues. Specify mass range. Only peptides with the masses within the specified mass range will be generated. You may also define an N-terminus or C-terminus modification, if nay, to a protein sequence. The number in box corresponds to the mass change from this modification. For example, a number '42.04' in the N-terminus box will give acetylation at the N-terminus. You may check the box for serine, threonine and tyrosine to include the possible phosphorylations on these amino acid residues.

4. Now you may calculate the specific degradation fragments for the parameters which you have specified, using 'Calculate Fragments' in 'Specific Degradation' menu. The number of generated fragments will be displayed on the screen. Because I only use a simple bubble sort routine, the sorting may be slow if the number of fragments is large.

5. Using the same menu, you may display the calculated results on the screen, or print them out. It is also possible to export the results to a file, so you may do further manipulations on them.

6. Using the 'Search' menu, you may also search a peptide sequence of certain mass or consensus sequence within a protein sequence. In this case, no specific degradation is considered. The program

searches all possible peptide sequences along the protein sequence. Specify the parameters in the 'Search Parameters' box on the screen.

7. You may also export a protein sequence to a file which can be converted to Macintosh file format, so you can utilize the functions in MacProMass software (by Terry Lee and Sunil Vemuri, Beckman Research Institute of the City of Hope, Durate, CA). See the 'HELP' menu for a detailed instruction.

Appendix VIII - An example of the program output of MSU MassMap v.1.1

Peptide Name: KRPSQRHGSKY, used as an example in Figure 5.1
1 KRPSQRHGSK Y
Number of Amino Acid = 11
Peptide Molecular Weight (MH^+) = 1344.5191
N-terminal modification = 0
C-terminal modification = 0
Cleave at: N-terminal of :
N-terminal of RK
Phosphorylation at: S,T,Y
mass range considered: 0 - 3000
total # of unphosphorylated fragments from this protein = 15
calculated # of unphosphorylated fragments within this mass range = 15
calculated # of phosphorylated fragments within this mass range = 22
calculated # of fragments within this mass range = 37

$[M+H]^+$, m/z	# of phosphorylation sites	peptide
147.20	0 p	1 - 1
175.21	0 p	2 - 2
182.20	0 p	11 - 11
262.18	1 p	11 - 11
303.38	0 p	1 - 2
428.47	0 p	7 - 10
487.54	0 p	3 - 6
508.45	1 p	7 - 10
567.52	1 p	3 - 6
591.64	0 p	7 - 11
643.72	0 p	2 - 6
671.62	1 p	7 - 11
723.70	1 p	2 - 6
751.60	2 p	7 - 11
771.90	0 p	1 - 6
851.88	1 p	1 - 6
896.98	0 p	3 - 10
976.96	1 p	3 - 10
1053.17	0 p	2 - 10
1056.94	2 p	3 - 10
1060.16	0 p	3 - 11
1133.15	1 p	2 - 10
1140.14	1 p	3 - 11
1181.34	0 p	1 - 10
1213.13	2 p	2 - 10

1216.35	0 p	2 - 11
1220.12	2 p	3 - 11
1261.32	1 p	1 - 10
1296.32	1 p	2 - 11
1300.10	3 p	3 - 11
1341.30	2 p	1 - 10
1344.52	0 p	1 - 11
1376.30	2 p	2 - 11
1424.50	1 p	1 - 11
1456.28	3 p	2 - 11
1504.48	2 p	1 - 11
1584.46	3 p	1 - 11

LIST OF REFERENCES

[1] Hillenkamp, F., Karas, M., Beavis, R. C., and Chait, B. T., *Anal. Chem.* **63,** 1193A-1202A (1991).

[2] Posthumus, M. A., Kistemaker, P. G., Meuzelaar, H. L. C., and Brauw, M. C., *Anal. Chem.* **50,** 985-991 (1978).

[3] Kupka, K.-D., Hillenkamp, F., and Schiller, C., *Advances in Mass Spectrometry,* Heyden & Sons., London, Vol. 8A, pp. 935-941 (1980)

[4] Karas, M. and Hillenkamp, F., *Anal. Chem.* **60,** 2301-2303 (1988).

[5] Karas, M., Bachmann, D., Bahr, U., and Hillenkamp, F., *Int. J. Mass Spectrom. Ion Processes* **78,** 53-68 (1987).

[6] Beavis, R. C. and Chait, B. T., *Rapid Commun. Mass Spectrom.* **3,** 432-435 (1989).

[7] Beavis, R. C., Chaudhary, T. and Chait, B. T., *Org. Mass Spectrom.* **27,** 156-158 (1992).

[8] Beavis, R. C. and Chait, B. T., *Rapid Commun. Mass Spectrom.* **3,** 436-439 (1989).

[9] Strupat, K., Karas, M., and Hillenkamp, F., *Int. J. Mass Spectrom. Ion Processes* **111,** 89-102 (1991).

[10] Spengler, B., Ying, P., Cotter, R. J., and Kan, L. S., *Rapid Commun. Mass Spectrom.* **4,** 99-102 (1990).

[11] Huth-Fehre, T., Gosine, J. N., Wu, K. J., and Becker, C. H., *Rapid Commun. Mass Spectrom.* **6,** 209-213 (1992).

[12] Nordhoff, E., Kirpekar, F., Karas, M., Cramer, R., Hahner, S., Hillenkamp, F., Kristiansen, K., Roepstorff, P., and Lezius, A., *Nucleic Acids Res.* **22,** 2460-2465 (1994).

[13] Tang, K., Allman, S. L., Chen, C. H., Chang, L. Y., and Schell, M., *Rapid Commun. Mass Spectrom.* **8,** 183-186 (1994).

[14] Schneider, K. and Chait, B. T., *Org. Mass Spectrom.* **28,** 1353-1361 (1993).

[15] Harvey, D. J., *Rapid Commun. Mass Spectrom.* **7,** 614-619 (1993).

[16] Stahl, B., Thurl, S., Zeng, J., Karas, M., Hillenkamp, F., Steup, M., and Sawatzki, G., *Anal. Biochem.* **223,** 218-226 (1994).

[17] Stahl, B., Steup, M., Karas, M., and Hillenkamp, F., *Anal. Chem.* **63,** 1463-1466 (1991).

[18] Huberty, M. C., Vath, J. E., Yu, W., and Martin, S. A., *Anal. Chem.* **65,** 2791-2800 (1993).

[19] Sutton, C. W., O'Neill, J. A., and Cottrell, J. S., *Anal. Biochem.* **218,** 34-46 (1994).

[20] Bahr, U., Deppe, A., Karas, M., and Hillenkamp, F., *Anal. Chem.* **64,** 2866-2869 (1992).

[21] Danis, P. O. and Karr, D. E., *Org. Mass Spectrom.* **28,** 923-925 (1993).

[22] Buerger, H. M., Mueller, H. M., Seebach, D., Boernsen, K. O., Schaer, M., and Widmer H. M., *Macromolecules* **26,** 4783-4790 (1993).

[23] Castro, J. A., Koster, C., and Wilkins, C., *Rapid Commun. Mass Spectrom.* **6,** 239-241 (1992).

[24] Castro, J. A., Koster, C., and Wilkins, C. L., *Anal. Chem.* **65,** 784-788 (1993).

[25] Wood, T. D., Schweikhard, L., and Marshall, A. G., *Anal. Chem.* **64,** 1461-1469 (1992).

[26] Ens, W., Mao, Y., Mayer, F., and Standing K. G., *Rapid Commun. Mass Spectrom.* **5,** 117-123 (1991).

[27] Karas, M., Bahr, U., and Giebmann, U., *Mass Spectrom. Rev.* **10,** 335-357 (1991).

[28] Demirev, P., Westman, A., Reimann, C. T., Hakansson, P., Barofsky, D., Sundqvist, B. U. R., Cheng, Y. D., Seibt, W., and Siegbahn, K., *Rapid Commun. Mass Spectrom.* **6,** 187-191 (1992).

[29] Spengler, B., Kirsch, D., Kaufmann, R., Karas, M., Hillenkamp, F., and Giebmann, U., *Rapid Commun. Mass Spectrom.* **4,** 301-305 (1990).

[30] Karas, M., Bahr, U., and Hillenkamp, F., *Int. J. Mass Spectrom. Ion Processes* **92,** 231-242 (1989).

[31] Karas, M., Bahr, U., Ingendoh, A., Nordhoff, E., Stahl, B., Strupat, K., and Hillenkamp, F., *Analytica Chimica Acta* **241,** 175-185 (1990).

[32] Beavis, R. C. and Chait, B. T., *Proc. Natl. Acad. Sci. USA* **87,** 6873-6877 (1990).

[33] Spengler, B., Kirsch, D., and Kaufmann, R., *J. Phys. Chem.* **96,** 9678-9684 (1992)

[34] Kaufmann, R., Kirsch, D., and Spengler, B., *Int. J. Mass Spectrom. Ion Processes* **131,** 355-385 (1994).

[35] Spengler, B. and Kaufmann, R., *Analusis* **20,** 91-101 (1992).

[36] Bahr, U., Karas, M., and Hillenkamp, F., *Fresenius J. Anal. Chem.* **348,** 783-791 (1994).

[37] Beavis, R. C. and Chait, B. T., *Anal. Chem.* **62,** 1836-1840 (1990).

[38] Beavis, R. C., *Org. Mass Spectrom.* **27,** 653-659 (1992).

[39] Brown, R. S. and Grifrich, N. L., *Rapid Commun. Mass Spectrom.* **6,** 690-696 (1992).

[40] Brown, R. S. and Grifrich, N. L., *Rapid Commun. Mass Spectrom.* **6,** 697-701 (1992).

[41] Kinsel, G. R., Grundwuermer, J. M., and Grotemeyer, J., *J. Am. Soc. Mass Spectrom.* **4,** 2-10 (1993).

[42] Vorm, O., Roepstorff, P., and Mann, M., *Anal. Chem.* **66,** 3281-3287 (1994).

[43] Schuerch, S., Schaer, M., Boernsen, K. O., and Schlunegger, U. P., *Biol. Mass Spectrom.* **23,** 695-700 (1994).

[44] Ingendoh, A., Karas, M., Hillenkamp, F., and Giessmann, U., *Int. J. Mass Spectrom. Ion Processes* **131,** 345-354 (1994).

[45] Grundwurmer, J. M., Bonisch, M., Kinsel, G. R., Grotemeyer, J., and Schlag, E. W., *Int. J. Mass Spectrom. Ion Processes* **131,** 139-148 (1994).

[46] Beavis, R. C. and Chait, B. T., *Rapid Commun. Mass Spectrom.* **3,** 233-237 (1989).

[47] Zaluzec, E. J., Gage, D. A., Allison, J., and Watson, J. T., *J. Am. Soc. Mass Spectrom.* **5,** 230-237 (1994).

[48] Barondes, S. H., Cooper, D. N. W., Gitt, M. A., and Leffler, H., *J. Biol. Chem.* **269,** 20807-20810 (1994).

[49] Dagher, S. F., Wang, J. L., and Patterson, R. J., *Proc. Natl. Acad. Sci. USA* ,in press (1995).

[50] Cowles, E. A., Agrwal, N., Anderson, R. L., and Wang, J. L., *J. Biol. Chem.* **265,** 17706-17712 (1990).

[51] Roff, C. F. and Wang, J. L., *J. Biol. Chem.* **258,** 10657-10663 (1983).

[52] Jia, S. and Wang, J. L., *J. Biol. Chem.* **263,** 6009-6011 (1988).

[53] Agrwal, N, Sun, Q, Wang, S. Y., and Wang, J. L., *J. Biol. Chem.* **268,** 14932-14939 (1993).

[54] O'Reilly, D. R., Miller, L. K., and Luckow, V. A., *Baculovirus Expression Vectors,* W. H. Freeman and Company, Oxford, England (1991).

[55] Agrwal, N, Wang, J. L., and Voss, P. G., *J. Biol. Chem.* **264,** 17236-17242 (1989).

[56] Sambrook, J., Fritsch, E. F., and Maniatis, T., *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).

[57] Laemmli, U. K., *Nature* **227,** 680-685 (1970).

[58] O'Farrell, P. H., *J. Biol. Chem.* **250,** 4007-4021 (1975).

[59] Millipore Corporation, *Investigator™ 2-D Electrophoresis System: Operating and Maintenance Manual,* Millipore Corporation, Bedford, MA (1993).

[60] Morrissey, J. H., *Anal. Biochem.* **117,** 307-310 (1981).