Fall 12-17-2011

Development of a novel electron-transfer secondary reaction matrix, characterization of the site–specificity of novel bilin-lyase, and Fundulus grandis protein expression investigation using mass spectrometry

Mohamed N. Boutaghou
Department of Chemistry-UNO, mboutag1@uno.edu

Development of a novel electron-transfer secondary reaction matrix, characterization of the site–specificity of novel bilin-lyase, and *Fundulus grandis* protein expression investigation using mass spectrometry

A Dissertation

Submitted to the Graduate Faculty of the
University of New Orleans
in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Chemistry

by

Mohamed Nazim Boutaghou

Bachelor of Science. University of Bordeaux 1, France, 2003
Master of Science. University of Bordeaux 1, France, 2005
Master of Management. IAE Bordeaux, France, 2007

December, 2011

Dedication

This work is dedicated to my parents and family who supported me throughout my college and graduate school and taught me that life is an eternal learning experience.

Acknowledgments

First and foremost, I would like to thank my advisor, Dr. Richard B. Cole, for allowing me to join his group and mentoring me throughout my graduate research. His insight and criticism helped overcome the challenges that arose in our projects. He taught me how to have a critical view of any results and how to make a scientific finding an irrefutable proof that it is occurring.

I had a chance to collaborate with Dr. Wendy M. Schluchter and Dr. Bernard B. Rees, and fortunately was able to interact with three advisors throughout my graduate research. They welcomed me to into their research projects and I was glad to contribute to their work. They contributed greatly in improving my knowledge of biology. I would like to thank them as well for proof-reading my chapters and for making valuable recommendations to improve chapters 3, 4 and 5. A special mention as well to Dr. Yang Cai for his valuable insights in database searching.

I would like to thank Dr. Avijit Biswas and Dr. Naga Abbaraju, who trusted me with their samples and their projects. I would like to thank as well my group members: Xiao Hua Liu, Nalaka Rannulu and Bing Guan for their support.

I also would like to thank my girlfriend, Jackie Mabry, and her family for accepting me as being a part of their family, without their support this dissertation would not have been finished.

I would also like to thank my mother, Ghania, my father, Mustapha and my sister Lila for their support. I would like to thank as well my sister Maya who has been a mentor throughout my college and graduate school.

Table of Contents

List of Figures

List of Tables

Abstract

The projects that are detailed in the coming pages take on a variety of subjects, but a common thread is that each employs matrix-assisted laser desorption/ionization (MALDI) mass spectrometry to solve a problem.

In the first project (Chapter 2), fundamental aspects of MALDI in-plume ionization are implicated in the introduction of a newly developed electron-transfer secondary ionization matrix. *9,10*-diphenylanthracene was successfully used for the analysis of various compounds that are not efficiently detected using common MALDI matrices. This improvement is likely related to a more efficient ionization through gas-phase electron-transfer as opposed to the commonly occurring gas-phase proton-transfer. Thermodynamic aspects underlying the gas-phase electron transfer involving *9,10*-diphenylanthracene were examined in detail. In particular, the possibility of gas-phase endothermic reactions was investigated.

In the second project, the site specificity of newly developed bilin-lyase enzymes (Chapter 3), Cpe Y/Z and CpeS was investigated. The results showed that CpeY/Z was the main bilin-lyase for phycorerythrobilin (PEB) attachment at Cysteine-82 of the C-phycoerythrin alpha subunit, whereas CpeS ligates PEB to Cysteine-80 of C-phycoerythrin beta subunits.

The third project (Chapter 4) arose from observations made during the investigation of PEB chromophore attachment. Reported in this chapter is a gas-phase trend noted for singly attached phycoerytrhobilin (PEB) that was developed into a means to distinguish between A-ring and A-ring plus D-ring double attachment of phycoerythrobilins. D-ring loss as a neutral was observed for all singly attached phycoerythrobilins and should not happen when double attahchment of PEB is occurring and the D-ring is covalently attached to a cysteine.

The fourth project (chapter 5) is related to protein identification using database searching. The protein expression pattern of *Fundulus grandis* was investigated in five different tissues. Reported in this chapter are the identifications of 405 proteins in muscle, liver, gill, heart and brain, with an emphasis on the evaluation of the false discovery rate. MALDI MS and MS/MS data were submitted to a "target-decoy" database search and the numbers of false positive identifications were deduced using two different methods. The impact of a very small number of false discoveries on the accuracy of our identifications was also discussed.

Key words: mass spectrometry, MALDI, Time of Flight, TOF/TOF, proteomics, database searching, bilin-lyase, electron-transfer secondary reactions, *9,10*-diphenyl anthracene, chlorophyll (a), retinol, C-phycoerythrin alpha subunit, C-phycoerythrin beta subunit, phycoerythrobilin, CpeY/Z, CpeS, tetrapyrrole, *Fundulus grandis*, false discovery rate, false-positive identification, "target-decoy" database search.

Chapter1

Introduction

In this introductory chapter, the basic concepts behind MALDI mass spectrometry will be presented. The formation of gas-phase ions, their analysis using a time-of-flight analyzer and the tandem mass spectrometry capabilities of the TOF/TOF analyzer will all be discussed. Mass spectrometry is an analytical technique that measures the mass-to-charge ratios of formed gas-phase ions. In-order to perform this task a mass spectrometer is composed of three-main structures: an ion-source, an analyzer and a computer for data processing. The function of the ion source is to efficiently bring an analyte in solution or in a solid mixture into the gas-phase while also inducing the creation of charged gas-phase molecules. Once the gas-phase ions formed, they are transferred to the analyzer through a series of electrostatic lenses that are positioned in separate compartments held at different pressures. To ensure that the analyzers function at high resolution, they must be held at very low-pressure in order to avoid collisions with background gases. Analyzers use properties of gas-phase ion physics to differentiate the various species of distinct m/z values.

1.1 MALDI Ionization

MALDI is a soft ionization technique that uses a matrix to promote formation of gas-phase ions of the analytes of interest. This ionization technique was introduced by Karas and Tanaka [1-3] and it became a method of choice for peptide and protein analysis. The matrix is a small molecule that forms solid crystals and absorbs efficiently at the laser wavelength. It is mixed with the analyte of interest in a variable matrix-to-analyte ratio (from a 100 to 50,000) and it acts as an intermediary in the process of sublimation of the analyte solid deposit and its ionization. The mechanism of MALDI ionization has been studied [4-6] and a concerted

mechanism has been proposed wherein the desorption/ablation event is followed by a two-step process: a primary ionization and a secondary ionization. The desorption / ablation process results from matrix absorption of laser photons and conversion of most of the photon energy into heat. The amount of heat present causes the matrix to vaporize and forces the analyte molecules to sublimate and desorb from the surface of the irradiated plate. It is at this earliest stage that primary ionization processes take place.

The primary ionization mechanisms are proposed to explain how ions are initially formed. It is possibly the most controversial part of MALDI ionization as no definitive proof has been given yet. Some of the most discussed models are known as "the preformed ion model" [7] and the "cluster model". [6] The preformed ion model, was built on the fact that positive ion peptide and protein analysis are typically carried out in an acidic medium, and therefore most of these analytes are already present as charged species before desorption. When they co-desorb along with the matrix analytes, they are brought into the gas-phase as charged species. The "cluster model" is a development of the preformed ion model. It states that analytes are present in the solid phase as clusters of ions with nearby counter ions. The ablative event needs to provide enough energy in order to separate the analyte from its counter ion in order to be detected. Even though the analyte and its counter ion can be separated, neutralization can later occur, which is why the detected analyte ions have been called "lucky survivors" [5]. Primary ions can be either matrix or analyte ions.

The secondary ionization mechanisms have much more of a consensus supporting them since it has been proven that local thermal equilibrium is reached in the MALDI plume. The MALDI plume is a region formed after laser induced desorption/ablation of the target (matrix plus analyte) mixture. The plume is characterized by a very high temperature (from 500 K to 1200 K) and a high density of particles that result from the desorption/ablation event. The density decreases as the plume expands into the vacuum. Primary reactions occur

in the denser part of the plume, whereas the secondary reactions occur in the expanding plume. Primary ions (matrix or analyte) react with neutrals in the expanding plume and form charged species. Charge transfer by the secondary reaction mechanism happens through proton, electron or cation transfer. These transfers can be either thermodynamically controlled or, less often, kinetically controlled. Physical properties of gas-phase analytes such as proton affinity, electron affinity and cation affinity are relevant constants when predicting the outcome of a secondary ionization reaction. More details on these reactions are provided in the next chapter, where a new electron transfer matrix is introduced.

1.2 Time-of-Flight Analyzers

Time-of-flight analyzers measure the time (t) an ion takes to travel a fixed distance (D), i.e., the length of the field-free drift region. When ions are formed in the source, they are accelerated using a high voltage to the same kinetic energy. This kinetic energy is proportional to the accelerating voltage V as shown in equation (1)

$$\tfrac{1}{2}\, m\, v^2 = eV \qquad (1)$$

$$and\ v = D / t \qquad (2)$$

Where, (m) is the ion molecular weight, (v) is the speed of the accelerated ion, (e) is the ion charge, (V) is the value of the accelerating voltage. By combining (1) and (2) we can derive equation (3) that defines the time of flight of an ion of mass ‘m’ and charge ‘e’ through the drift tube:

$$t = (m / 2eV)^{\frac{1}{2}} D \qquad (3)$$

Therefore, by measuring the time an ion takes to reach the detector, one can deduce the value its mass-to-charge ratio. Time-of-flight analyzers have been widely coupled with MALDI

ionization sources, because MALDI ionization uses a pulsed laser (not a continuous laser) operated at a certain frequency as opposed to a continuous source emission. Each laser pulse is therefore used as a starting time, t_0 at which to begin measurement of ion travel time. Since all of the formed ions have the same kinetic energy, according to equation (1), smaller ions will have a higher speed than heavier ions. Therefore arrival of lighter ions will precede the arrival of heavier ions.

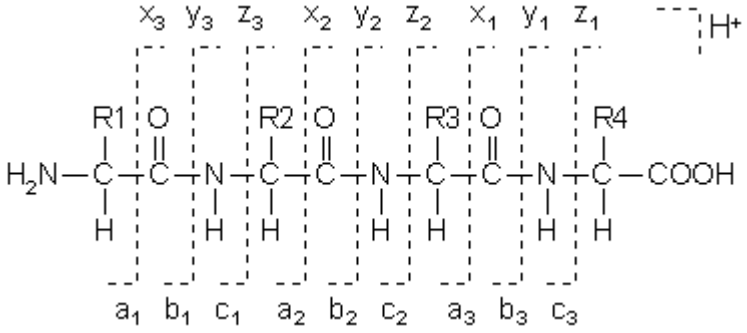
One of the limitations of TOF analyzers is the problem of the initial kinetic energy spread. This limitation arises from because ions having an initial kinetic energy are being accelerated to what is supposed to be a uniform kinetic energy. However, the initial energy will be added to the kinetic energy provided by the accelerating voltage. The additional energy, if positive in the direction of the flight path, will cause the ion to have a shorter travel time than other identical ions that have no initial energy prior to acceleration. This results in a broader peak for the detected ion and a lowering of resolution. In order to address this issue, an "ion mirror" for kinetic energy focusing was introduced [8,9] and instruments that employ this technology are known as a "reflectron" time of flight analyzers, (RTOFs). The idea behind it is to add a retarding electric field at the end of the drift region, which will slows down ions and reverses their direction. Ions with aberrantly high kinetic energies will penetrate deeper into the electric field and will have a longer travel path. As a result, faster ions of a given m/z value will travel on slightly longer paths than slower ions of the same m/z value. This causes ions of the same m/z value to arrive at the detector at the same time (i.e., they are energy focused) regardless of their initial kinetic energy prior to acceleration.

1.3 Tandem mass spectrometry on a TOF analyzer

The 4800 MALDI TOF/TOF (AB Sciex, Toronto, Canada) mass spectrometer used in our experiments is equipped with second stage accelerating lenses which is equivalent to having a second TOF analyzer in series with the first [10]. The second source is located after a collision cell where decomposition occurs, thus tandem mass spectrometry experiments are enabled. Ions that are formed in the first ion source are accelerated to 6-8 keV by a first stage of lenses. A timed ion selector deflects away all unwanted ions, but then lets the ion of interest pass. This selected precursor of interest is decelerated to 1 or 2 keV before it enters the collision cell. Upon entering the collision cell, the precursor ion collides with neutral gas molecules. These collisions cause the precursor to fragment. The product ions enter a second source where they are reaccelerated and refocused before being detected.

Among the most widely investigated precursor ions for such tandem MS experiments are protonated and multiply protonated peptides. Peptide MS and MS/MS analyses have been widely used in order to identify proteins and/ or identify protein post-translational modifications. [11, 12] Collision induced decomposition of peptides will generate product ions that result from cleavages of the on the amino-acid backbone (figure 1.1) or side chains. The most common cleavage occurs on the amide bond (CO-NH). The nomenclature describing the product ion takes into account the localization of the positive charge after the cleavage. If the positive charge remains on the N-terminal side, the product ion is called a *b* ion. If the charge is located on the C-terminal fragment, the product ion is known as a *y* ion. Other fragment ions can result from cleavage of the $CH(R_n)$-CO bond (*a* and *x* ions) and the NH-$CH(R_n)$ bond (*c* and *z* ions) [13]. Peptide mass measurement and fragmentation are the main experiments that are used to identify proteins.

(A)



(B)

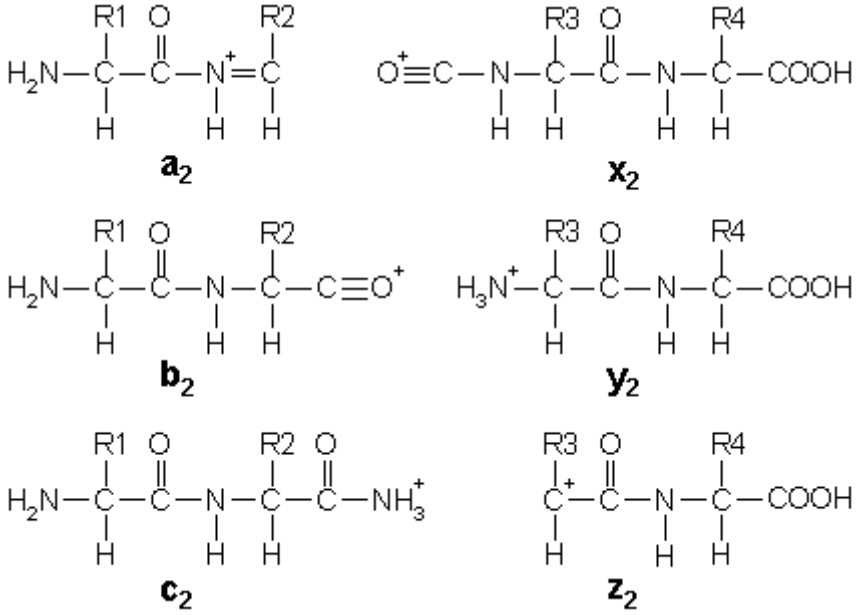


Figure 1.1. (A) Peptide CID fragment nomenclature, *b* and *y* ions are the most common fragments. (B) Structure of the different fragments. *a*, *b* and *c* fragment ions will retain the charge on the N-terminal of the precursor peptide. *x*, *y* and *z* ions will have the charge retained on the C-terminal of the peptide precursor. The numbering in subscript is related to the number of amino acids contained in the fragments [14].

1.4 Protein Identification and Database Searching

MALDI MS and MS/MS spectra of proteolytic peptides are used to identify unknown proteins. The identification process is called "protein database searching" or "peptide to spectrum match" [15, 16] and it proceeds on two levels. The first level is known as "peptide mass fingerprinting" [14, 17]. MALDI MS spectra are compared to *in-silico* generated MS peaks originating from proteolysis of proteins of known sequences contained in the database. Once a match between peaks contained in the experimental dataset and the theoretical one occurs, a second level of identification will take place. The second level of identification is known as "peptide fragment fingerprinting". Experimental product ions of the peptide of interest are matched against *in-silico* generated product ions of a peptide of known sequence and originating from the database. A protein is identified when several MALDI MS peaks are matched with a peptide of a protein belonging to the database. The accuracy increases when these peptides are matched using their masses (PMF) and their sequences (PFF). These searches are performed using software algorithms. Several database searching tools are available: Sequest™ [18] ProteinProspector™ [19] and MASCOT™.

MASCOT scoring scheme is based on a scoring algorithm introduced by Papin et al [20] called MOWSE: MOlecular Weight SEarch. The MOWSE search program uses peak lists derived from acquired mass spectra (Fragment Molecular weights, FMWs), along with a user specified error tolerance (Usually 1-2 Da. Up to 5 Da) and matches them with masses of in-silico generated peptides (DBMws = Database Molecular weights) using a specific enzyme (cleavage rules of each specific enzyme are applied). According to the search program, a "hit" is registered when:

$$DBMw - tolerance-1 < FMw < DBMw + tolerance+1$$

The scoring strategy will use each FMw submitted for the search and calculate a frequency of occurrence of each FMw in proteins belonging to the database. A distribution frequency is assigned for each FMw. All the distribution frequencies of the "hits" are multiplied in order to obtain the P_N which is the distribution frequency score of each N matching fragment. The final score is calculated as:

$$\text{Score} = 50000 / (P_N \times H)$$

H is the molecular weight of the matched protein.

The investigators tested their search algorithm by submitting a hundred test proteins to tryptic digestion. The search was successful at identifying 99% of the protein correctly using 5 or less peptides (mean of 3.6). This figure could be lowered to 2.7 peptides average/protein if the tolerance in mass-to-charge ratio measurement is lowered to +/- 1Da. Based on the empirical results and the scoring algorithm, MOWSE favors identifications that consider longer peptide lengths and instruments that have higher mass accuracies. MOWSE was the first search algorithm to emphasize that the relative abundance of a peptide of a given length resulting from a proteolytic digest will depend upon its length and the length of the protein.

MASCOT [21] scoring employs a probability based algorithm (MOWSE algorithm uses only occurring frequencies). Mascot will calculate the probability (P) for a matched sequence to be a random match. The displayed score is calculated as follows:

$$\text{Score} = -10 \times \text{Log}(P/k)$$

k = number of sequences contained in the database

The other development that Mascot offers is the ability to use MS spectra (PMF) and associated MS/MS spectra (PFF = Peptide Fragment Fingerprint) in order to evaluate the probability of a hit to be random. The higher the score, the lower will be the probability of a match being random. The peptide with the lowest probability of being random is assigned as the best match.

1.5 References

[1] Karas, M.; Bachman, D.; Bahr, U.; Hillenkamp, F. *Int. J. Mass Spectrom. Ion Processes* **1987**, *78,* 53-68

[2] Karas, M.; Hillenkamp. F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10,000 Daltons. *Anaytical Chemistry.* **1988**, *60,* 2299-2301

[3] Tanaka, K,; Waki, H; Ido, Y.; akita, S.; Yoshida, Y.; Yoshida, T. *Rapid Commun. Mass Spectrom.* **1988**, *2,* 151-153

[4] Dreisewerd, K. The desorption process in MALDI. *Chemical Reviews.* **2003**, *103,* 395-426

[5] Karas, M.; Kruger, R. Ion formation in MALDI: The cluster ionization mechanism. *Chemical Reviews* **2003**, *103,* 441-452

[6] Knochenmuss, R.; Zenobi, R. MALDI ionization: the role of in-plume processes. *Cehmical Reviews* **2003**, *103,* 441-452

[7] Richard Knochenmuss. MALDI Ionization Mechanisms: an Overview. From *electrospray and MALDI mass spectrometry: Fundamentals, instrumentation, Practicalities and Biological Applications*, Second edition, Edited by Richard B. Cole. 2010, pages 149-175. (c) 2010 John Wiley & Sons.

[8] Robert J. Cotter. MALDI Ionization Mechanisms: an Overview. From *electrospray and MALDI mass spectrometry: Fundamentals, instrumentation, Practicalities and Biological Applications*, Second edition, Edited by Richard B. Cole. pages 345-364. (c) 2010 John Wiley & Sons

[9] Mamyrin, B. A.; Karataev, V. I.; Shmikk, D. V.; Zagulin, V. A. *Soviet Physics JETP* **1973**, 37, 45

[10] Medzihradszky, K. F., Campbell, J. M.; Baldwin, M. A.; Falick, A.M.; Juhasz, P.; Vestal, M. L.; Burlingame, A.L.; The characteristics of peptide collision-induced dissociation using a high performance MALDI-TOF/TOF tandem mass spectrometer. *Analytical Chemistry* **2000**, 72, 552–558

[11] Mann, M.; Jensen, O, N. Proteomics analysis of post-translational modifications. *Nature Biotechnology*. **2003**, 21, 255-261

[12] Aebersold, R.; Goodlett, D. R. Mass Spectrometry in Proteomics. **2001**, 101, 269-295

[13] Biemann, K., Appendix 5. Nomenclature for peptide fragment ions (positive ions). *Methods in Enzymology.* **1990**, 193, 886-887

[14] MASCOT help page: http://www.matrixscience.com/help/fragmentation_help.html

[15] Marcotte, E. M. How do shotgun proteomics algorithms identify proteins? *Nature Biotechnology* **2007**, 25, 755-757

[16] Dworzanski, J. P., Deshpande, S. V., Chen, R., Jabbour, R. E., Snyder, A. P., Wick, C. H., Li, L.; Mass Spectrometry-Based Proteomics Combined with bioinformatics Tools for Bacterial Classification. *Journal of Proteome Research* **2006**, 5, 76-87

[17] Thiede, B., Hohenwater, W., Krah, A., Mattow, J., Schmid, M., Schmidt, F., Jungblut, P. R. Peptide mass fingerprinting. *Methods* **2005**, 35, 237-247

[18] Eng, J. K.,Mc Cormack, A. L., Yates, J. R. An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. *Journal of the American Society for Mass Spectrometry* **1994**, 5, 976-989

[19] Clauser, K. R., baker, P., Burlingame, A. L. Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS MS and database searching. *Analytical Chemistry* **1999**, 71, 2871-2872

[20] Papin, D. J. C., Hojrup, P., Bleasby, A. J.; Rapid identification of proteins by peptide-mass fingerprinting. Current Biology. 1993, 3, 327-332

[21] Perkins, D. N., Pappin, D. J. C., Creasy, D. M., Cottrell, J.S.; Probability-based protein identification by searching sequence database using mass spectrometry data. Electrophoresis. 1999, 20, 3551-3567

Chapter 2

Introducing *9,10*-diphenyanthracene as a matrix for MALDI electron transfer secondary reactions

2.1 Abstract

The most common secondary-ionization mechanism in positive ion MALDI involves a gas-phase proton transfer reaction to ionize the analyte. Peptides and proteins are molecules that have basic (and acidic) sites that make them susceptible to gas-phase proton transfer. However, non-polar, aprotic compounds that lack basic sites are more difficult to protonate and creating charged forms of this type of analyte can pose a problem when conventional MALDI matrices are employed. In this case, forming a radical molecular ion through electron-transfer is a viable alternative and certain matrices may facilitate the process. In this work, we investigate the performance of a newly developed electron-transfer secondary reaction matrix: *9,10*-diphenyl anthracene. The use of *9,10*-diphenylanthracene as matrix for MALDI analysis has been tested using several model compounds. It appears to promote ionization through electron-transfer in a highly efficient manner as compared to other potential matrices. Thermodynamic aspects of the observed electron transfers in secondary-ionization reactions were also considered, as was the possibility of kinetically controlled / endothermic, electron-transfer reactions in the MALDI plume.

2.2 Introduction

Matrix Assisted Laser Desorption Ionization (MALDI) is a sensitive ionization technique that uses a matrix in order to desorb and ionize analytes of interest in preparation for mass spectrometric analysis [1- 3]. The matrix compound must absorb efficiently at the laser wavelength, it must also co-crystallize and co-desorb with the analyte upon irradiation. Once in the gas-phase, the primary matrix ion reacts with neutral analyte molecules by secondary reactions that, in the positive ion mode, will involve proton, cation and electron transfers.

Most of the matrices developed for positive ion MALDI mass spectrometry are acidic in nature, and are thus well suited to undergo proton transfer. The most widely used matrices that fall into this category are low molecular weight acidic aromatic compounds such as: α-cyano-4-hydroxycinnamic acid (CHCA) [4] and 3,5-dimethoxybenzoic acid (sinapinic acid) [5], which have been extensively used for peptide and protein analyses. Analyte protonation using such matrices is controlled by gas-phase proton transfer thermodynamics which state that a proton transfer is favored from compounds with high acidity (or low proton affinity (PA)) to compounds with higher PA [6]. Gas-phase metal ion transfers have been described as well, and this strategy has been used mainly for analytes that have proton affinities lower than those of available matrices. Thus, when proton transfer is not favorable, metal cations may still attach. Silver trifluoracetate has been successfully used in this way for MALDI analysis of polystyrene polymers [7], and lithium was used successfully for MALDI analyses of various lipid classes [8, 9].

Electron transfer secondary reactions involving analytes are less commonly reported in MALDI-MS. Although all common matrices are capable of forming radicals, they are not always suitable for accepting/donating electrons as compared to competing and often more

favorable proton transfer. Electron transfer reactions are favored secondary ionization mechanisms for low polarity polymers that lack, or hold few, polar functional groups. When reactions are thermodynamically controlled, an ion whose neutral form has a higher ionization potential (IP) will capture an electron from a compound with a lower IP upon collision [10-12]. If the experiment is performed in the negative mode, the electron affinity becomes the relevant parameter. In the thermalized plume, electron transfer reactions are thought to yield the thermodynamically favored products [13]. Early work on electron transfer matrices described the use of various non-polar matrices: acenaphthene, perylene, terthiophene and anthracene for the analysis of metallocenes [11]. *Trans*-2-[3-(4-*t*-butyl-phenyl)-2-methyl-2-propenylidene]malononitrile (DCTB) matrix has been successfully applied to MALDI generation of fluorofullerene anions [14], and analysis of polythiophene [15]. Tetrathiafulvalene (TTF) has been used as a matrix for MALDI analysis of pigments in the negative mode [16].

9,10-Diphenylanthracene (*9,10*-DPA) is a material used in Organic Light-Emitting Diodes (OLEDs). It was reported as an electron transfer reagent for use in Electron Transfer Dissociation (ETD) reactions [17]. In those studies, *9,10*-DPA$^{\cdot}$ was found to transfer an electron efficiently upon collision with multiply charged peptides, although proton transfer from the peptide was also possible. Based on its optical properties and its ability to transfer electrons, we have investigated the use of *9,10*-diphenyl anthracene as an electron transfer secondary reaction matrix. *9,10*-DPA's ability to generate radical cations and anions was investigated against a variety of compounds and its behavior was compared to those of traditional MALDI matrices.

2.3 Materials and Methods

Chemical reagents *9,10*-diphenylanthracene, *9,10*-dibromoanthracene, azulene, zinc protoporphyrin, and buckminsterfullerene, C_{60}, were purchased from Alfa Aesar (Ward Hill, MA). Anthracene, 9-methylanthracene, fluoranthene, *trans*-stilbene, pyrene, chlorophyll a, vitamin A, vitamin K1 (phylloquinone), vitamin K2 (penaquinone), and 4-chromanol were purchased from Sigma-Aldrich (St. Louis, MO). Acetone and dichloromethane were also purchased from Sigma-Aldrich.

Mass Spectrometry MALDI-MS and MALDI-MS/MS experiments were performed on an Applied Biosystems (Foster City, CA) / MDS Sciex (Concord, Ontario) 4800 MALDI TOF/TOF spectrometer. Mass spectral acquisitions were obtained in the reflectron mode using a Nd:YAG laser operated at 355 nm. The dried droplet spotting technique was used; an equal volume of matrix and test analyte were mixed in an Eppendorff tube and 0.5 ul of the mixture was spotted on a MALDI 384-well plate and air-dried. All mass spectra were acquired in the reflectron mode with an average of 400 laser shots / spectrum. Mass spectral processing and peak area quantification were performed using the 4000 Series Explorer peak list editor.

Absorption Absorption spectra were obtained using a SpectraMax M2 spectrophotometer, employing a Xenon flash lamp as the light source.

Sample preparation Test matrices were dissolved in acetone, except for *9,10*-dibromo anthracene which was dissolved in dichloromethane. All test matrices and test compounds were prepared on the day of the analysis. Matrix and analyte solutions were prepared so as to achieve a spot concentration of 5-15 nmol and an analyte: matrix ratio of 1:100.

2.4 Results and Discussion

In order to provide a thorough assessment of the utility of *9,10*-diphenyl anthracene for producing radical molecular ions, its performance was juxtaposed with various aromatic compounds that have previously been used as electron transfer secondary reaction matrices: terthiophene, anthracene, and pyrene. In addition to these three, other aromatic compounds that could potentially transfer an electron to the analyte, i.e., fluoranthene, azulene were also tested. All of the compounds selected for evaluation had two common characteristics: 1) they were cyclic aromatic systems and 2) they lacked acidic (and basic) sites, thus they would not promote proton transfer reactions. We also included in our comparison traditional MALDI matrices as "bench marks". Table 1 compiles the list of compounds, their molecular weights and their ionization potentials.

	Compound name	Molecular wt.	IP (eV)	Laser energy (a.u) threshold for $M^{+\cdot}$ appearance	Molar Absorptivity (ε) at 355nm (L. $mol^{-1}.cm^{-1}$)
Test Compounds as electron transfer secondary matrices	*9,10*-DPA (*9,10*-Diphenyl Anthracene	330.14	7.0 [18]	2200	7220
	9,10 -DBA (*9,10*-DibromoAnthracene)	333.90	7.5 [18]	2800	4561
	Anthracene	178.07	7.8 [4, 13]	3200	3150
	Fluoranthene	202.07	7.9 [19]	3300	3255
	Terthiophene	247.97	7.3 [4]	2000	28085
	Pyren	202.07	7.4 [20]	3000	462
Usual MALDI matrices	α-CHCA	189.04	8.5 [6]	N/A	
	Sinapinic Acid	224.07	7.7 [6]	N/A	
	2,5 DHB	154.03	8.1 [21]	N/A	
	Dithranol	226.06	6.9 [6]	N/A	

Table 2.1. Ionization Potentials of test compounds as electron transfer secondary reaction matrices. Laser desorption energy threshold (a.u) for detection of the radical molecular cation, $M^{+\cdot}$ of test matrices (12 nmoles). Molar absorptivities of test matrices were calculated, terthiophene and 9,10-DPA are the most absorbing matrices at the laser wavelength, λ=355nm.

We first tested the ability of the matrix to absorb photons at the laser wavelength, λ=355 nm. Absorption spectra were acquired in the UV-Visible and molar absorptivities (ε) were calculated (Table 2.1). Equimolar solutions of test matrices were prepared, spotted and subjected to laser desorption ionization (LDI). The laser was ramped up from the lowest possible fluence until the radical molecular cation peak was detected; laser energy appearance thresholds (the lowest laser energy yielding a detectable signal) are reported in Table 1. There is a clear correlation between the compound's ability to absorb laser photons (as indicated by ε) and the threshold of appearance of its radical molecular cation. Terthiophene and *9,10*-DPA (fig. 2.1), the most absorbing test matrices, were detected at the lowest laser threshold.

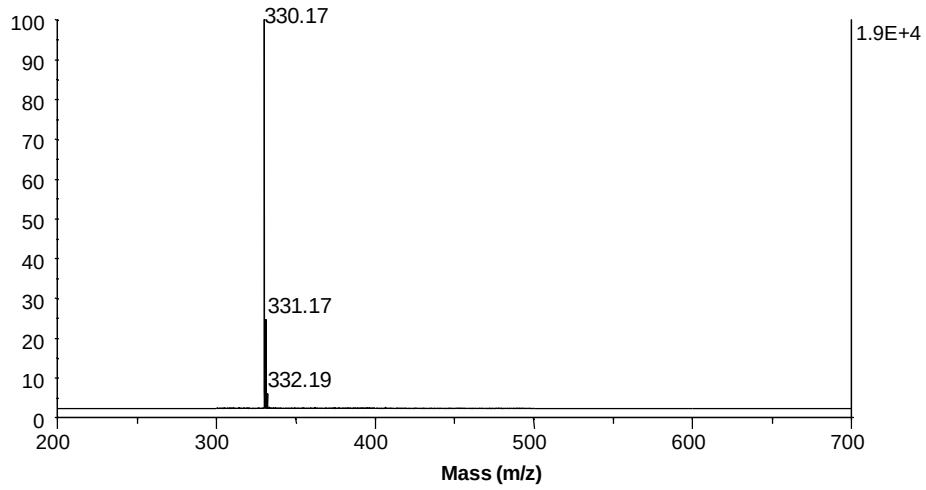


Figure 2.1. Mass spectrum of a laser desorption experiment of *9*,10-DPA. A strong peak at m/z 330 is detected. The isotopic pattern of the detected radical (A, 100%; A+1, 30%; A+2, 4%) matched perfectly with the theoretical isotopic pattern: A, 100%; A+1, 29%; A+2, 4%.

In a follow up experiment, test matrices were mixed with a series of compounds in order to investigate their abilities to undergo secondary electron transfer reactions. All of the analytes were first run without a matrix in order to assess their abilities to form radical cations or to form MH^+ upon laser desorption ionization (LDI, no matrix). Chlorophyll (a) was the first test analyte; it contains an extensively conjugated tetrapyrrole center, which is known for stabilizing radical cations. Chlorophyll (a) and (b) have been previously analyzed using Fast Atom Bombardment (FAB) [22, 23] and laser desorption mass spectrometry (LDMS) [24, 25]. The radical cation, as well as the protonated molecule, were readily observed in FAB using 3-nitrobenzyl alcohol as the liquid matrix [18]. Laser desorption MS analysis of chlorophyll (a) using our 355 nm Nd:YAG laser did not yield any molecular ion peak (radical or protonated). A small peak corresponding to $M^{+\cdot}$ can be seen at high laser fluences, but such laser power concurrently causes extensive fragmentation. Chlorophyll (a) was mixed individually with each test matrix in our library and MALDI MS analysis was then performed. The results are summarized in figure 2.3. A major fragment at m/z 614 was detected in laser desorption experiments as well as in all experiments employing matrices. This fragment has been assigned as $[M\text{-}C_{20}H_{38}]^{+\cdot}$ and it is proposed to result from the cleavage of the phytyl moiety of chlorophyll (a) (figure 2.2) after proton transfer via a 6-membered ring intermediate. When mixed with 4-CHCA and DHB, MALDI analysis of chlorophyll (a) shows a peak at m/z 893, which corresponds to protonated chlorophyll (a), MH^+. Major fragmentation resulting in the demetalation of chlorophyll yields $[M\text{-}Mg]^{+\cdot}$ at m/z 870 and $[MH - Mg]^+$ at m/z 871. It has been reported that acidic matrices demetalate metalloporphyrins in the gas phase [26]. However, these fragments were minor when using sinapinic and dithranol matrices. Both of these latter matrices were able to generate a radical molecular ion, which is detected at m/z 892 with some protonated molecules appearing at m/z 893. When comparing the performance of all of the matrices, *9,10*-diphenyl anthracene

enabled the detection of the radical molecular cation in the highest S/N and without causing extensive demetalation. The ionization energy of Chlorophyll (a) has been determined to be approximately 6.1 eV [27]. *9,10*- DPA had the lowest ionization energy (IE) amongst all the tested matrices, at 7.04 eV. This causes the electron transfer from chlorophyll (a) to be less exothermic to *9,10*-$DPA^{+\cdot}$ as compared to radical cation forms of the other matrices, resulting in less fragmentation, which increases the intensity of the $M^{+\cdot}$ peak.

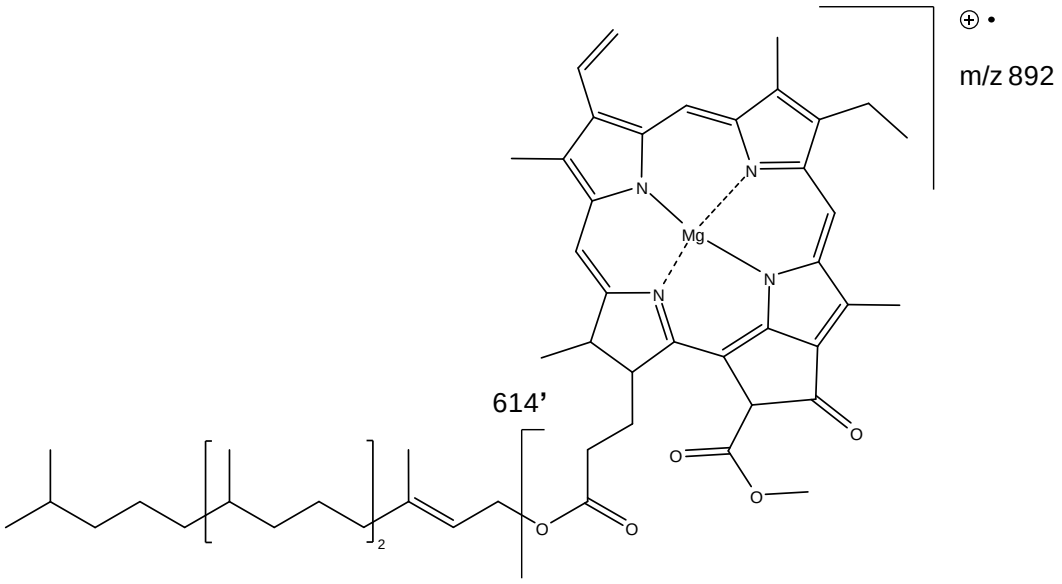


Figure 2.2. Structure of chlorophyll (a). The peak at m/z 614, which is not shown, is a major fragment detected upon MALDI analysis of chlorophyll (a). It results from the cleavage of the phytyl group. The tick mark on the right of 614 indicates that a proton has been transferred to the formed ion upon cleavage.

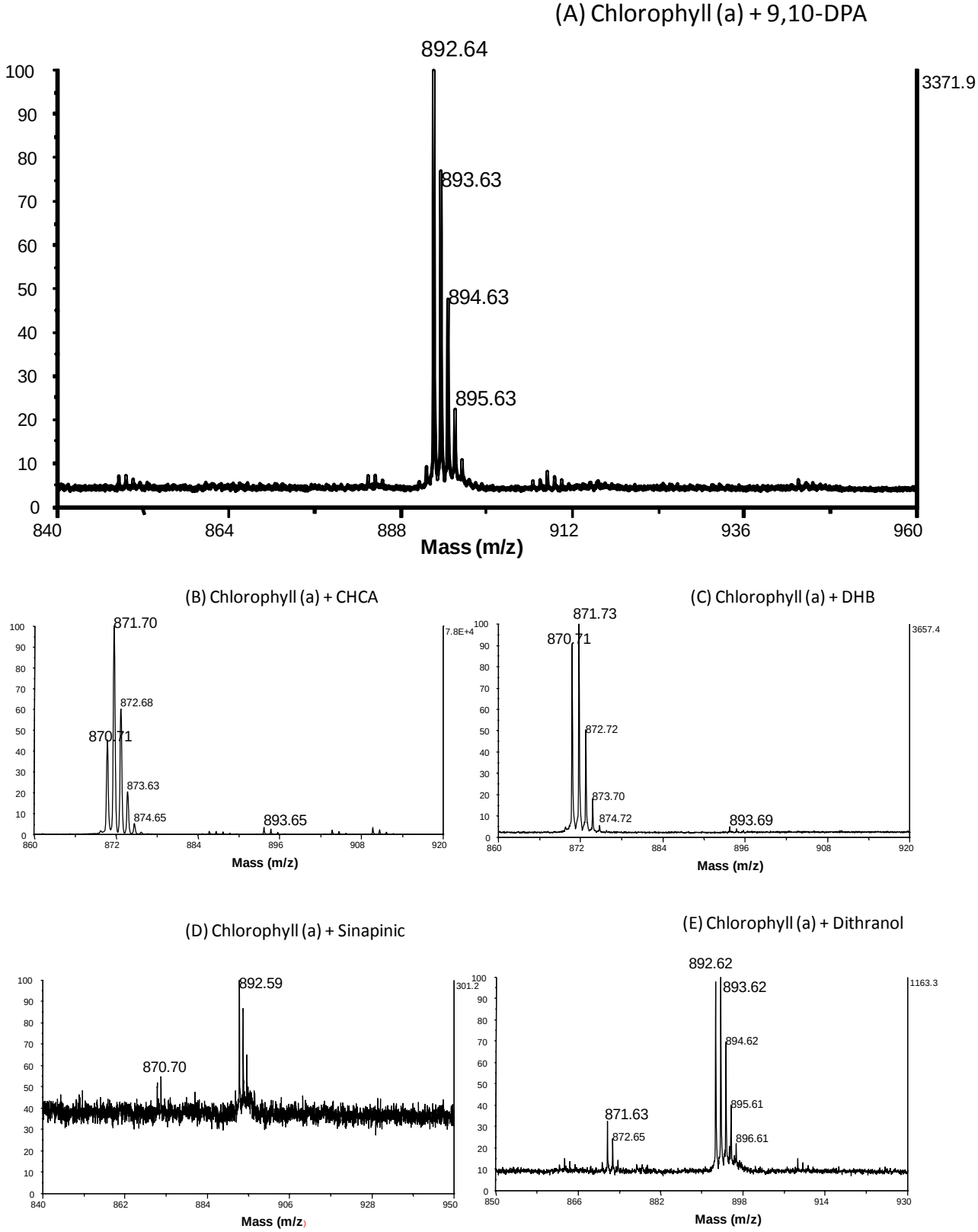


Figure 2.3. Chlorophyll (a) analysis using, (A) *9,10*-DPA ,(B) 4-CHCA, (C) *2,5*-DHB. (D) Sinapinic Acid, and (E) Dithranol. The peak detected at m/z 892 represents the radical molecular ion. The peak detected at m/z 893 represents the protonated molecular ion $[M+H]^+$. The peak at m/z 870 was identified as the demetalated radical cation $[M-Mg+2H]^{+\cdot}$ and m/z 871 was identified as the protonated demetalated chlorophyll $[M+H-Mg+2H]^+$. Demetalation of chlorophyll (a) has been reported when mixed with acidic matrices. *9,10*-DPA has the lowest ionization potential, which makes the electron transfer reaction softer and causes little fragmentation.

Next, retinol (Vitamin A) was employed to evaluate *9,10*-DPA vs the other test matrices. Previously published work has established that, during laser desorption (no matrix), retinol readily forms a radical cation [28]. However, when subjected to laser desorption MS experiments using our MALDI TOF/TOF, neither the radical molecular ion (m/z 286) nor the protonated molecule (m/z 287) was detected from retinol. When analyzed by MALDI using regular matrices, CHCA (figure 2.4.B), dithranol (figure 2.4.E), DHB (figure 2.4.C), sinapinic acid (figure 2.4.D), retinol tends to form $[M+H]^+$ (m/z 287) with dominant water loss, i.e., $[M-H_20]^+$ (m/z 269) is the base peak in Figures 2.4.B-2.4.E. Water loss occurs as well from the radical molecular cation, mechanism of such mechanism is described in figure 2.5. By contrast, when using *9,10*-DPA (figure 2.4.A), the radical molecular ion, $M^{+\cdot}$ (m/z 286) becomes the base peak and the peak at m/z 269 becomes minor. Pyrene promotes the formation of $M^{+\cdot}$ as well, but the peak at m/z 269 is still a major peak The remaining test compounds did not promote formation of the radical molecular ion and the resulting spectra displayed little to no signal.

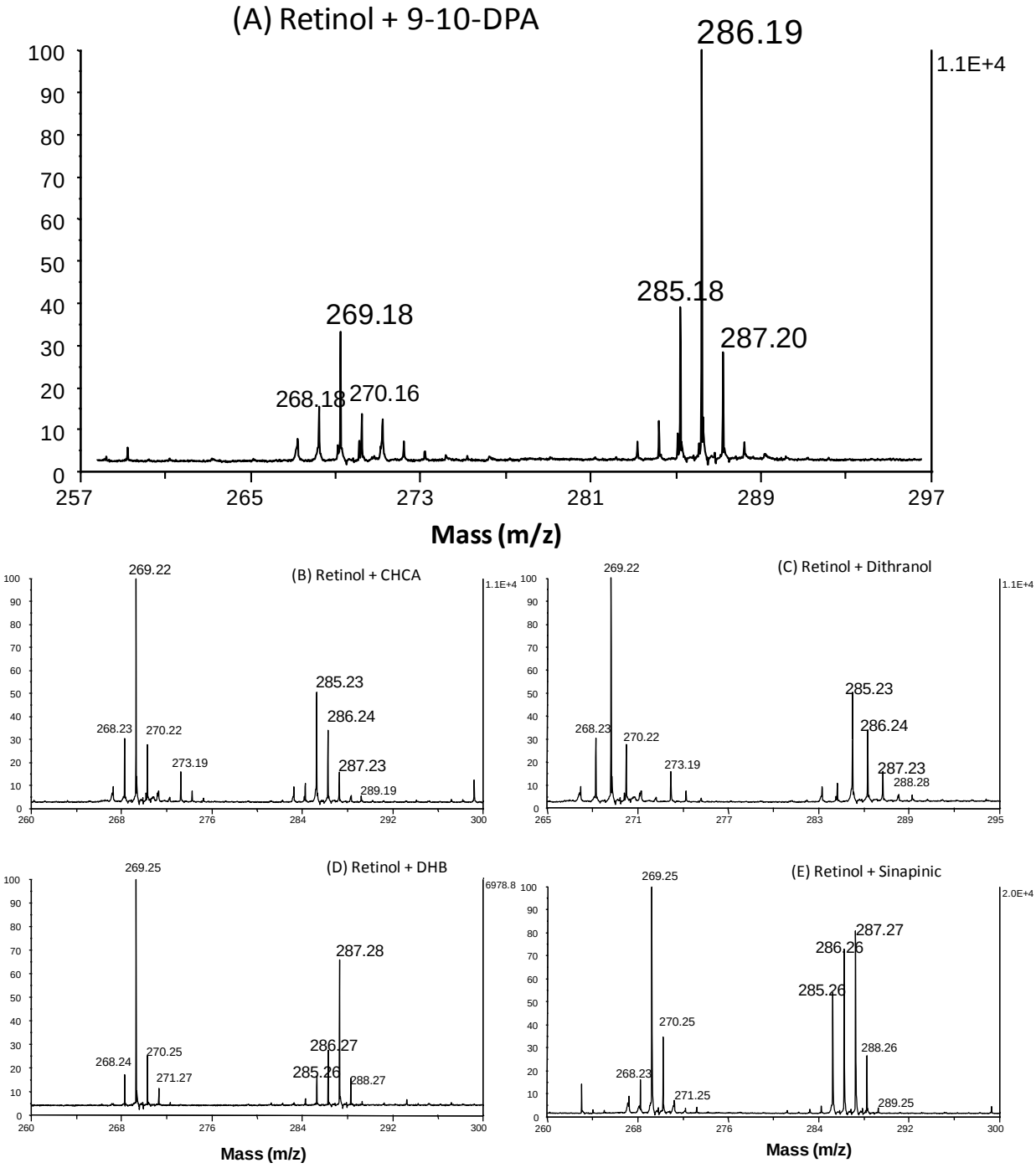


Figure 2.4. MALDI MS mass spectra of retinol analysis using different matrices. (A) *9, 10*-DPA, (B) CHCA, (C) Dithranol, (D) 2,5-DHB and (E) Sinapinic acid. Radical molecular ion detected at m/z 286 is the base peak using *9,10*-DPA.

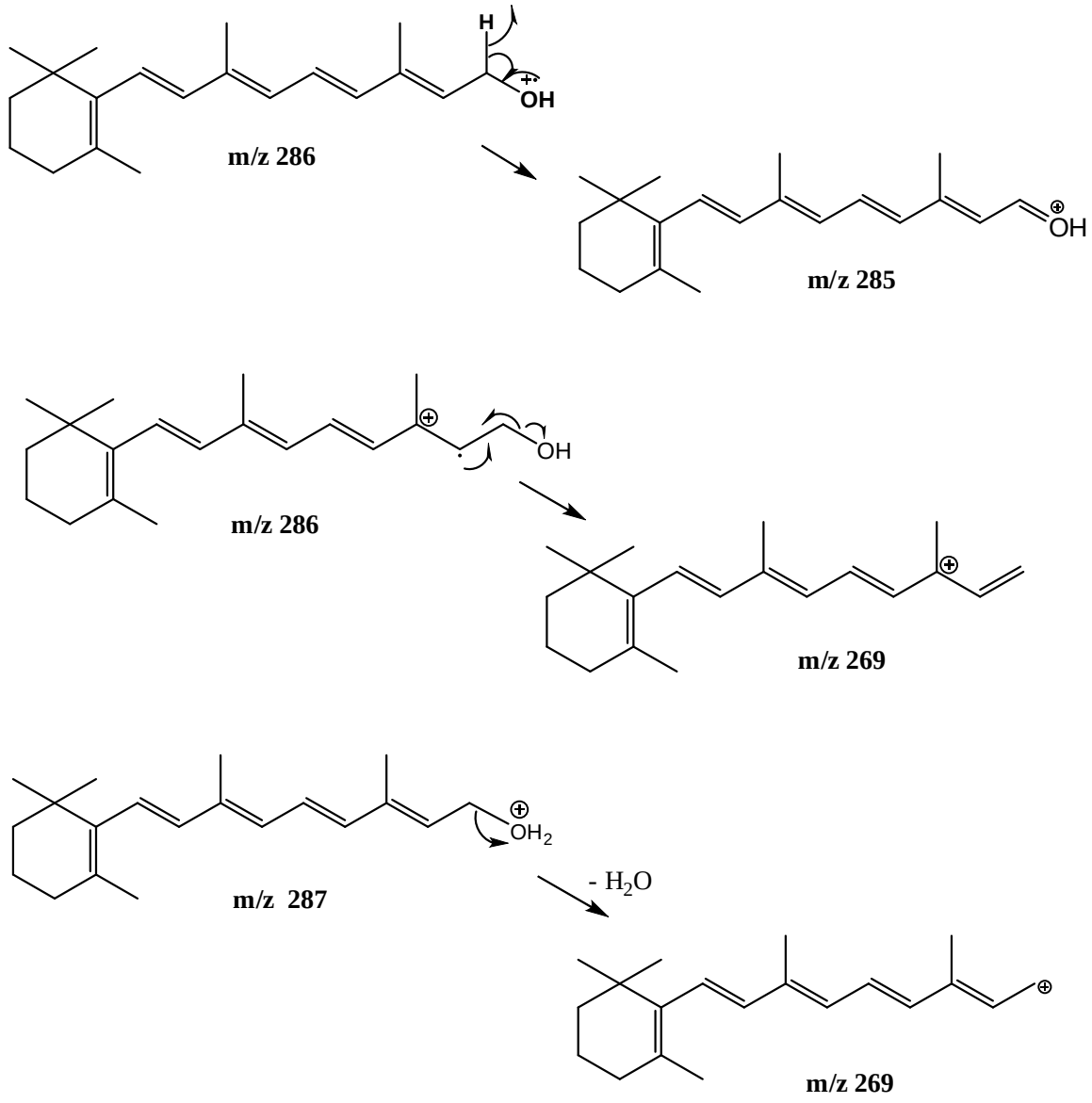


Figure 2.5. Structure of retinol and mechanism of formation of formed fragments. The peak at m/z 286 was the base peak when analyzed using *9,10*-DPA.

Several other compounds were employed to test the ability of the matrix *9,10*-DPA to promote electron transfer secondary reactions. Results are summarized in table 2.2.

Test compound	Molecular wt.	Laser Desorption Ionization	MALDI analysis using 9,10-DPA as a matrix
Chlorophyll (a)	892	--	++++
Retinol (Vit. A)	286	--	++++
Zinc Protoporphyrin	624	++++	++++
Menaquinone (MK-4)	444	+	++++
Phylloqinone (K1)	450	+++	++++
Meso-tetrakis (pentafluorophenyl)porphyrin	974	++++	++++
Cholesterol	386	--	--
4-Chromanol	150	--	--
fullerene MS(+)	720	++++	++
Fullerene MS(-)	720	+	++++

Table 2.2. Summary of test analyte analyses with and without *9,10*-DPA. When adding *9,10*-DPA as a matrix, the increase in the radical molecular cation ($M^{+\cdot}$) signal intensity was noticeable for analytes such as chlorophyll (a), retinol, menaquinone, phylloquinone and (fullerene in the negative mode). No ($M^{+\cdot}$) was detected for cholesterol and 4-chromanol.

Menaquinone (vitamin K2, MK-4), a blood clotting agent, was analyzed using *9,10*-DPA and its radical molecular cation peak intensity grew two-fold compared to the peak resulting from Laser Desorption MS. Buckminsterfullerene (C_{60}) was also tested as an analyte for *9,10*-DPA, but in the negative mode. In the absence of *9,10*-DPA, LDI experiments on fullerene in the negative mode detected a low intensity radical molecular anion. When mixed with *9,10*-DPA, the radical anion peak intensity increased ten-fold (figure 2.6). In all previous experiments, carried out in the positive mode, *9,10*-DPA was the

primary matrix ion $M^{+\cdot}$ and as this radical cation, it was available as an electron acceptor. In the negative mode, *9,10*-DPA becomes an electron-donor, and as such, it was successful at transferring an electron to the fullerene molecule. These results indicate that *9,10*-DPA has the potential to serve as a dual matrix for both positive and negative mode analyses.

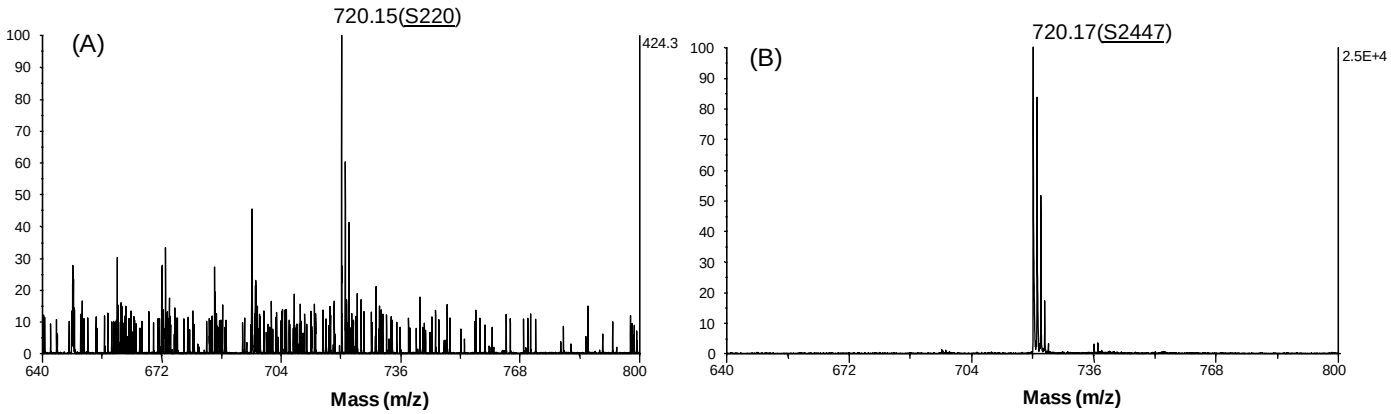


Figure 2.6. (A) LDI mass spectrum of a fullerene in the negative mode. (B) MALDI mass spectrum in the negative mode, of fullerene using *9,10*-DPA. The intensity of the ($M^{+\cdot}$**)** peak increased tenfold when adding *9,10*-DPA.

Thermodynamic evaluation of the electron transfer process using DPA

The direction of gas-phase electron transfer can be predicted based on the relative ionization energies of the interacting compounds in the plume. A neutral analyte may become ionized to form a radical molecular cation if one of its electrons can be transferred to $M^{+\cdot}$ of the matrix. This reaction is favored thermodynamically if the ionization energy of the analyte is lower than the ionization energy of the matrix. The IE of Chlorophyll (a) has been determined to be in the range of [4.5-6 eV] [29-31]. When considering the higher IE of *9,10*-DPA (7.04 eV) [32], electron transfer from chlorophyll to *9,10*-$DPA^{+\cdot}$ is thus exothermic. Various ionization energies of retinol have been published with values ranging from 6.95 eV [32] to 7.3 eV [33] to 7,4 eV [34]. MALDI analysis of retinol mixed with *9,10*-DPA as a matrix, resulted in detection of an intense radical molecular cation (m/z 286). According to these published values, electron transfer from retinol to *9,10*-$DPA^{+\cdot}$ is approximately thermoneutral or even slightly endothermic. Of course, endothermic reactions are not favored, and McCarley et al [10] and Hoteling et al [12] both stressed that such endothermic reactions should not occur in the plume, although in each publication, some results suggested that endothermic reactions may be possible. It is our view that a slight endothermicity might be overcome by a modest collision energy in the plume. We decided to investigate thermodynamic aspects of electron transfer reactions between compounds of known IP values. We chose to monitor peak areas of compounds in binary and ternary mixtures as a function of laser fluence, using *9,10*-DPA as the primary matrix.

A solution of pyrene (IP= 7.4 eV) and *9,10*-DPA (7.04 eV) was mixed at an equimolar ratio and submitted to a laser power ramp experiment. Peak areas for each compound are reported as a function of laser fluence (figure 2.7).

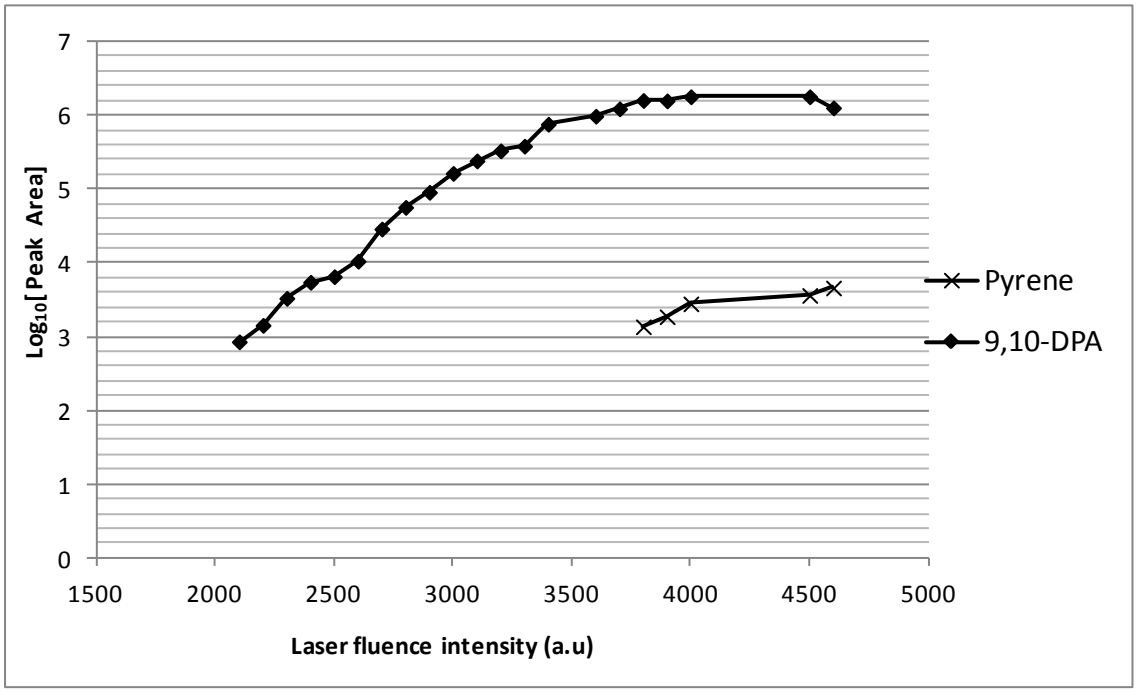


Figure 2.7. Plot of Log_{10}[peak area] as a function of laser fluence intensity. When laser fluence reaches 3800, a peak at m/z 202 belonging to the radical molecular cation of pyrene was detected. Since the IP of pyren (7.4 eV) is higher than *9,10*-DPA (7.04 eV) the detection of pyrene radical cation can be interpreted as resulting from an endothermic electron transfer reaction.

No peaks for the radical molecular cation of pyrene ($\mathbf{p}^{+\cdot}$) were detected at the low laser energy. This result is not unexpected given that pyrene has a higher ionization potential, and even if ($\mathbf{p}^{+\cdot}$) were to be formed, it would rapidly neutralize after accepting an electron from neutral *9,10*-DPA. However, at a laser fluence value of 3800, a peak corresponding to ($\mathbf{p}^{+\cdot}$) was indeed observed, and its intensity increased with increasing laser energy. The formation of the pyrene radical cation ($\mathbf{p}^{+\cdot}$) could be due to two reasons: 1) neutral pyrene absorbs laser photons and forms a radical cation as a primary ionization reaction, or 2) neutral pyrene endothermically transfers an electron to *9,10*-$DPA^{+\cdot}$. According to the molar absorptivities reported in table 2.1, pyrene absorbs very little at 355 nm, which renders doubt about the possibility of ($\mathbf{p}^{+\cdot}$) formation from direct laser absorption and increases the plausibility that this radical was formed via endothermic electron transfer. In order to gain

more information, a third compound, *9,10*-dibromoanthracene (*9,10*-DBA) (IP=7.54 eV [34]) was selected and mixed in a binary solution with *9,10*-DPA and separately in a ternary solution with *9,10*-DPA and pyrene. The rationale behind performing this experiment was that, if there is an endothermic electron transfer between *9,10*-DPA and *9,10*-DBA, adding pyrene, which has a lower IP than *9,10*-DBA, should create a competitive endothermic reaction with *9,10*-DPA, which will lower the amount of *9,10*-$DBA^{+\cdot}$ formed. Figure 2.8, shows a plot of *9,10*-$DBA^{+\cdot}$, m/z 333, peak area versus laser fluence for the binary and ternary mixtures.

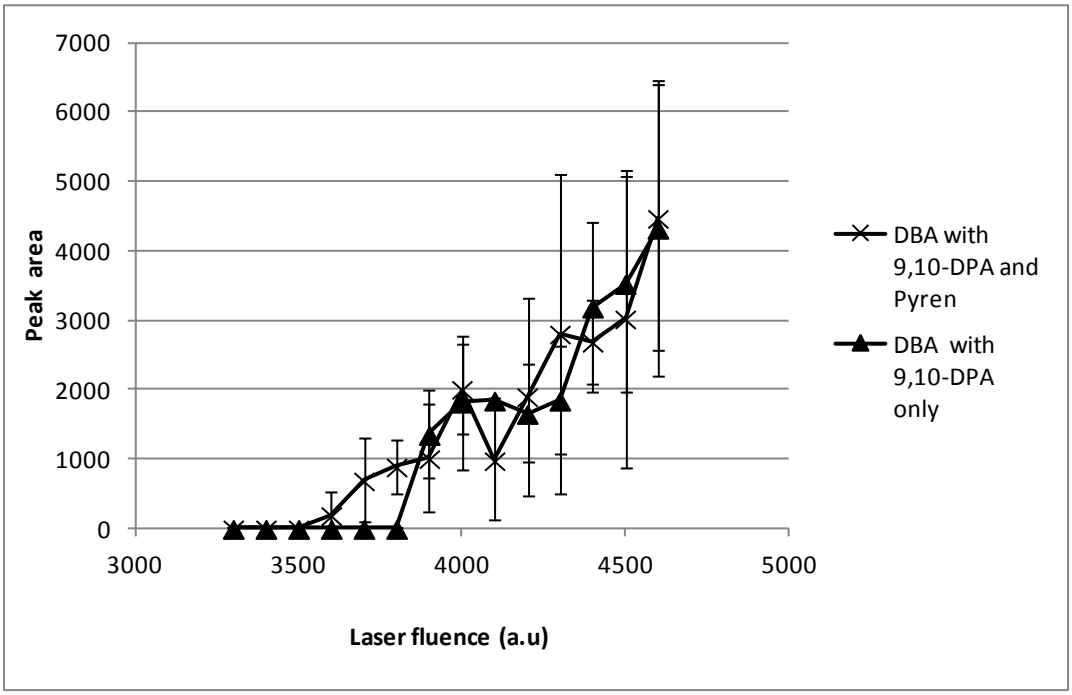
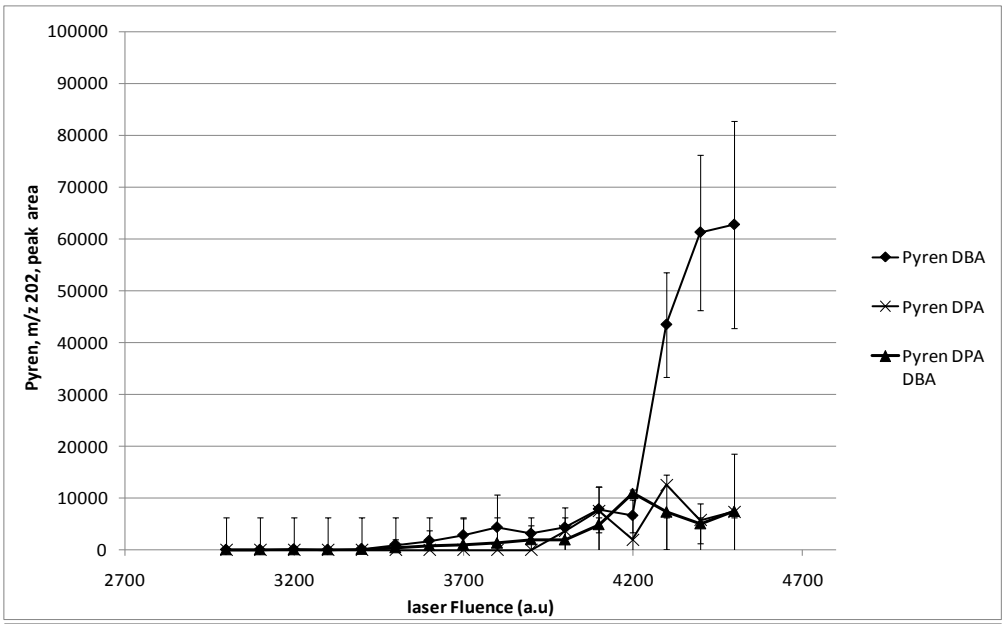


Figure 2.8. [*9,10*-$DBA^{+\cdot}$] peak area with respect to laser fluence. The peak area was monitored in a binary mixture with *9,10*-DPA (data points with a triangle shape) and in a ternary mixture, with *9,10*-DPA and Pyren (data points with a cross shape). The amount of *9,10*-DBA radical molecular ion detected is not affected by the solution composition. Its formation is interpreted as resulting from direct photoionization.

The 2 plots are superimposable for the most part, and they do not show a significant change in *9,10*-DBA$^{+\cdot}$ formation when adding a competing analyte for the endothermic reaction. This result demonstrates that *9,10*-DBA$^{+\cdot}$ formation is unchanged in the presence of pyrene, and this ion is probably formed from direct photo-ionization from the laser. This latter deduction is supported by *9,10*-DBA absorptivity (table 2.1) that show that *9,10*-DBA is more highly absorbing than pyrene at the laser wavelength of 355nm.

In a follow up experiment, we decided to monitor the pyrene ($\mathbf{p}^{+\cdot}$) peak from binary solutions composed of pyrene with *9,10*-DBA or *9,10*-DPA and in a ternary mixture with all three compounds. Figure 2.10.A shows a plot of the peak area of ($\mathbf{p}^{+\cdot}$), m/z 202, with respect to laser fluence for the three different mixtures. When mixed with *9,10*-DBA, pyrene radical molecular cation is detected at the lowest laser threshold when compared to the two other solutions (figure 2.10.B). The average peak area of ($\mathbf{p}^{+\cdot}$) is higher when mixed with *9,10*-DBA. This is probably due to the fact that ($\mathbf{p}^{+\cdot}$) formation results from an exothermic electron transfer to *9,10*-DBA$^{+\cdot}$.When mixed in a binary solution with *9,10*-DPA, the radical molecular cation appears at the highest laser fluence and its average peak intensity is very close to the peak intensity detected when mixed in a ternary solution. When compared with *9,10*-DBA and *9,10*-DPA, pyrene has an intermediate IP and should undergo an exothermic electron transfer to *9,10*-DBA$^{+\cdot}$, but could afterwards be neutralized by the exothermic transfer of an electron from *9,10*-DPA. The amount of formed ($\mathbf{p}^{+\cdot}$) from pyrene mixed with *9,10*-DPA only and when mixed with both *9,10*-DBA and *9,10*-DPA is very close, leading us to propose that the detected radical molecular cation of pyrene is formed in each of these experiments via the same mechanism: direct pyrene absorption of laser photons (even though its absorbtivity is poor at this wavelength). The possibility of an endothermic electron cannot be established based on our results.

(A)



(B)

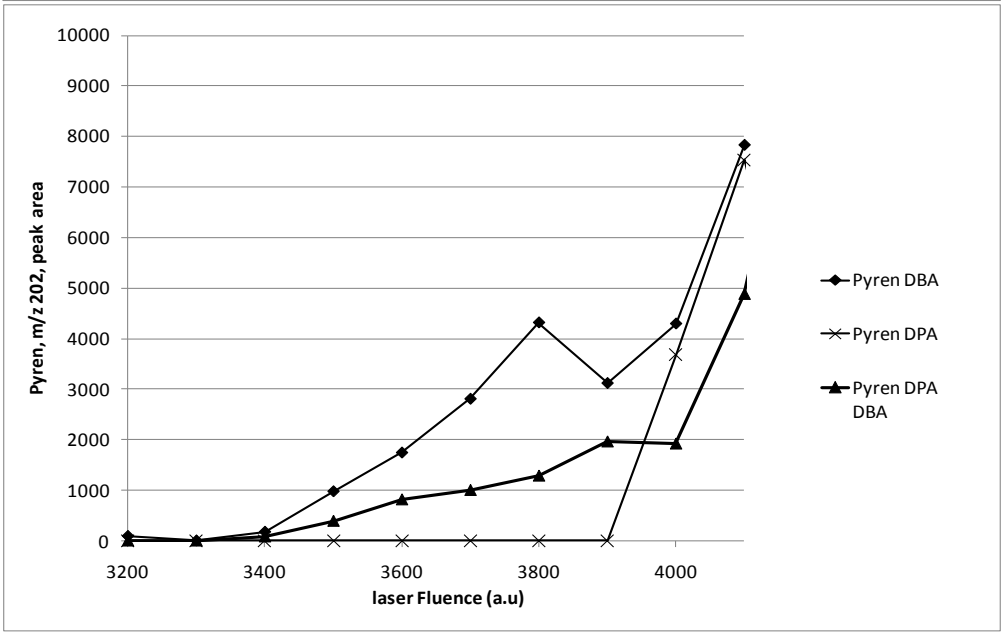


Figure 2.9. Pyren, m/z 202, peak area with respect to laser fluence. Pyren peak area was monitored when mixed in a binary mixture with *9,10*-DPA (cross data point), *9,10*-DBA (diamond data points) and in a ternary mixture with *9,10*-DPA and *9,10*-DBA (triangle data point). Pyren has the lowest detection threshold when mixed with *9,10*-DBA, and the highest when mixed with *9,10*-DPA. Pyren was detected at the highest intensity when resulting from an exothermic electron transfer reaction from *9,10*-DBA. Pyren radical molecular cation peak intensity does not vary much when mixed with *9,10*-DPA only and both *9,10*-DPA and *9,10*-DBA, leading us to think that its formation results from direct absorption of laser photons.

Instances of endothermic electron (or even proton) transfer reaction are seldom mentioned in the MALDI literature. However, a report of kinetically-controlled gas-phase proton transfer reactions [6] established that there was a better chance for a MALDI in-plume proton transfer reaction to be kinetically controlled at lower, rather than high, laser fluences. The rationale for this deduction is that under such conditions, the amount of desorbed matrix and analyte desorbed is low which causes very little in-plume collisions. The plume expands fast and does not allow secondary reactions to achieve completion [35]. Pyrene radical molecular ion was detected at higher laser fluences, which, according to this rationale, would decrease the chances of observing an endothermic reaction. But MALDI in-plume electron transfer reactions can be kinetically controlled. Hoteling et al [12] presented results that showed that when mixing a matrix with analytes of much lower IP (over 1.5 eV in differential between the matrix IP (8.3 eV) and the analyte IP (6.5 eV)) and at high matrix to analyte ratio, the relative intensity of the analyte did not reflect the ratios of a thermodynamically controlled reaction, but rather, a kinetically controlled one.

2.5 Conclusions

9,10-diphenyl anthracene was introduced and characterized as a new matrix for secondary electron-transfer reactions. Radical molecular ions of Chlorophyll (a), retinol and several other compounds were successfully detected in MALDI experiments employing this new matrix. *9,10*-DPA displays a strong laser photon absorbance and an intense radical molecular cation can be detected at low laser fluences. Electron transfer in the gas-phase, under thermodynamic control, will require the analyte to have a lower ionization potential than *9,10*-DPA (7.04 eV). Because the IP of *9,10*-DPA is so low when electron transfer occurs, it causes fewer fragmentations than any of the other matrices investigated. The thermodynamics of electron transfer with analytes of higher IP than *9,10*-DPA were also investigated, but evidence of formation of radical molecular cations via endothermic electron transfer was not conclusive. Appearance of radical cations from compounds with higher neutral IPs than DPA was more likely due to independent laser photon absorption than an endothermic electron-transfer reaction.

2.6 References

[1] Karas, M.; Bachman, D.; Bahr, U.; Hillenkamp, F. *Int. J. Mass Spectrom. Ion Processes* **1987**, *78,* 53-68

[2] Karas, M.; Hillenkamp. F. *Anal. Chem.* **1988**, *60,* 2299

[3] Tanaka, K,; Waki, H; Ido, Y.; akita, S.; Yoshida, Y.; Yoshida, T. *Rapid Commun. Mass Spectrom.* **1988**, *2,* 151-153

[4] Beavis, R.; Chaudhary, T.; Chait, B. *Org. Mass Spectrom.* **1992**, *27*, 156-158

[5] Beavis, R.: Chait, B. *Rapid Commun. Mass Spectrom.* **1989**, *3*, 432-435

[6] Breuker, K.; Knochenmuss, R.; Zhang, J.; Stortelder, A.; Zenobi, R. *Int. J. Mass Spectrom.* **2003**, *226,* 211-222

[7] Bahr, U.; Depee, A.; Karas, M.; Hillenkamp, F.; Giessman, U. *Anal. Chem.* **1992**, *64,* 2866-2869

[8] Suzuki, Y.,; Suzuki, M.; Ito, E.; Goto-Inoue, N.; Miseki, K.; Lida, J.; Yamazaki, Y.; Yamada, M.; Suzuki, A. *J. Biochem.* **2006**, *139,* 771-777

[9] Trimpin, S.; Clemmer, D. E.; McEwen, C. N. *J. Am. Soc. Mass Spectrom.* **2007**, *18,* 1967-1972

[10] Mc Carley, T.D.; Mc Carley, R. L.; Limbach, P. A. *Anal. Chem.* **1998**, *70,* 4376-4379

[11] Macha, S. F.; Mc Carley, T. D.; Limbach, P. A. *Anal. Chim. Acta* **1999**, *397,* 235-245

[12] Hoteling, A. J.; Nichols, W. F.; Giesen, D. J.; Lenhard, J. R.; Knochenmuss, R.; *Eur. J. Mass Spectrom.* **2006**, *12,* 345-358

[13] Knochenmuss, R.; Zhigilei, L. V. *J. Mass Spectrom.* **2010**, *45,* 333-345

[14] Streletskii, A. V.; Ioffe, I. N.; Kotsiris, S. G.; Barrow, M. P.; Drewello, T.; Strauss, S. H.; Boltalina, O. V. *J. Phys. Chem.* A **2005**, *109,* 714-719

[15] De Winter, J.; Deshayes, G.; Boon, F.; Coulembier, O.; Dubois, P.; Gerbaux, P. *J. Mass Spectrom.* **2011**, *46,* 237-246

[16] Akasawa, D.; Chen, L. C.; Hiraoka, K. *J. Mass Spectrom.* **2008**, *43,* 1494-1501

[17] Coon, J.J.; Syka, J. E. P.; Schwartz, J. C.; Shabanowitz, J.; Hunt, D. F. *Int. J. Mass Spectrom.* **2004**, *236,* 33-42

[18] Masnovi, J. M.; Seddon, E. A.; Kochi, J. K. *Can. J. Chem*. **1984**, *62,* 2552-2559

[19] Ling, Y.; Lifshitz, C. *J. Phys. Chem*. **1995**, *99,* 11074-11080

[20] Hager, J.W.; Wallace, S.C.; *Anal. Chem*. **1988**, *60,* 5-10

[21] Karbach, V.; Knochenmuss, R. *Rapid. Com. Mass Spectrom*. **1998**, *12,* 968-974

[22] Brereton, R. G.; Bazzaz, M. B.; Santikam, S.; Williams, D.H. *Tetrahedron Lett.***1983**, *24,* 5775-5778

[23] Grese, R. P.; Cerny, R. L.; Gross, M. L.; Senge, M. *J. Am. Soc. Mass Spectrom.* **1991**, *1,* 72-84

[24] Tabet, J.-C.; Jablonski, M.; Cotter, R. J.; Hunt, J. E. *Int. J. Mass Spectrom Ion Processes.***1985**, *65,* 105-107

[25] Dale, M. J.; Costello, K.F.; Jones, A.C.; Langrige-Smith, P. R. R. *J. Mass Spectrom.* **1996**, *31,* 590-601

[26] Srinivisan, N.; Haney, C. A.; Lindsey, J. S.; Zhang, W.; Chait, B.T. *J. Porphyrins Phthalocyanines* **1999**, *3,* 283-291

[27] From NIST CHEM webbook: Nakato, Y.; Chiyoda, T.; Tsubomura, H.; *Bull. Chem. Soc. Jpn.* **1974**, *47,* 3001.

[28] Wingerath, T.; Kirsh, D.; Spengler, B.; Stahl, W. *Anal. Biochem.* **1999**, *272,* 232-242

[29] Bernas, A.; Grand, D.; Hautecloque, S.; Myasoedova, T. *Chem. Phys. Letters* **1984**, *104,* 105-108

[30] Kashiwagi, H.; Hirota, F.; Nagashima, U.; Takada, T. *Int. J. Quant. Chem.* **1986**, *30,* 311-326

[31] Pandey, A.; Datta, S, N. *J. Phys. Chem. B.* **2005**, *109,* 9066-9072

[32] Katsumata, S.; Ikehata, N. *J. Electron Spec. Rel. Phenom.* **2000**, *107*, 139-145

[33] Smith, G.J. *Photochem. Photobiol.* **1983**, *38,* 119-121

[34] Guaratini, T.; Vessecchi, R. L.; Lavarda, F. C.; Maia Campos; P.; Naal, Z.; Gates, P. J.; Lopes, N. P.; *Analyst*, **2004**, *129,* 1223-1226

[35] Knochenmuss, R. MALDI Ionization Mechanisms: an Overview. In *Electrospray and MALDI Mass Spectrometry: fundamentals, Instrumentation, Practicalities, and Biological Applications, Second Edition*; Cole, R. B., Ed.; John Wiley & Sons, 2010, Chapter 5, pp 149-184

Chapter 3

Localization of the specific site of enzyme-mediated covalent binding of phycoerythrobinlin (PEB) to C-phycoerythrin alpha and beta subunit

3.1 Abstract

This work investigates the enzyme-catalyzed attachment of a specific bilin (phycoerythrobilin, PEB) to the C-phycoerythrin alpha subunit (CpeA) and beta subunit (CpeB) of the cyanobacterium *Fremyella Diplosiphon* strain UTEX 481. The goal was to determine the site-specificity of CpeYZ- and CpeS- mediated bilin attachment. An off-line LC–MALDI-TOF/TOF method was developed to locate the exact site(s) of PEB covalent attachment to the host proteins (CpeA and CpeB). For PEB attachment to CpeA catalyzed by CpeY and CpeZ, a major HPLC peak was detected when tryptic peptides were separated on a reversed-phase HPLC column. MALDI MS of the HPLC fraction detected a peak of interest at m/z 935. The MS/MS spectrum of the PEB-CpeA complex precursor at m/z 935 contained two peaks of interest: one at m/z 587 that was attributed to protonated PEB, and one at m/z 349 that corresponds to a protonated tripeptide containing a cysteine to which the PEB was covalently bonded. This site of attachment was unequivocally deduced to be cysteine 82. For the CpeS-mediated CpeB-PEB attachment, one major tryptic peptide was collected after RP-HPLC separation. MALDI MS analysis of the fraction resulted in detection of a peak at m/z 1250 whose product ion (tandem mass spectrum) contained a peak at m/z 587, as well as a peak at m/z 664. M/z 1250 was assigned as a protonated peptide containing a cysteine at position 80, to which the PEB was covalently bonded. These findings identified the PEB attachment sites for the CpeYZ enzyme as Cys-82 on CpeA and the CpeS enzyme as Cys-80 on CpeB.

3.2 Introduction

Cyanobacteria, plants and algae are living organisms that utilize “light” as a primary source of energy. Photosynthetic organisms, such as cyanobacteria, contain two types of photosystems: photosystem I and photosystem II. Large antennas absorb the light and transfer the resulting excitation energy into the photosystems thus enabling electron transport. Phycobiliproteins are the major light absorbing pigment in Photosystem II. They are organized into a bigger complex, known as the Phycobilisome and have been isolated from red algae and cyanobacteria [1]. Phycobiliproteins are multimeric complexes that arise from the assembly of several subunits: alpha, beta and sometimes gamma. Each subunit will undergo several post-translational modifications in order to reach its native conformation. One of these modifications is the attachment of linear tetrapyrrole molecules to highly conserved sites on each subunit. These tetrapyrrole molecules are known as “phycobilins” or “bilins”, and their attachment is catalyzed by bilin lyases.

Free bilins, as well as denatured biliproteins, are unsuitable photoreceptors because they absorb light poorly and their excited states are very short-lived, leading to conversion of their excitation energy into heat. On the other hand, the phycobiliprotein subunit-bilin complex will display much improved photophysical properties. The light absorption of the chromophore is increased by one order of magnitude, and the excited lifetimes by four orders of magnitude, the combination of which, results in an excellent photoreceptor unit [2]. This improvement is related to the conformation of the bilin pigment. Attached PCB has an extended planar structure, proven by X-ray crystallography, and they display a much improved absorption in the visible. Free bilins fluoresce very weakly because of the loss of the excitation energy due to favored radiationless relaxation pathways. Once bound, the

linear and sustained rigid structures of the chromophores will minimize those effects and will cause the native bilins to fluoresce with higher quantum efficiencies [3-5].

Phycocyanobilin was the first bilin structure to be reported [6,7], followed shortly thereafter by the structure of PEB [8]. The structures of several additional phycobilins have been described: phycoviobilin, PVB, phycourobilin, PUB, mesobiliverdin and dihydrobiliverdin as well as others [9]. Their structures were determined primarily by ^{1}H NMR [6-8]. Phycobilins have a similar structure to heme and heme derived-bilins, and the possibility that they are products of the tetrapyrrole biosynthetic pathway was investigated. A study showed phycocyanobilin incorporated ^{14}C when bacteria were fed radiolabeled δ-aminolevulinic Acid [^{14}C-ALA] [10]. ALA is the universal precursor for all the tetrapyrrole biosynthetic pathways and originates, in cyanobacteria, from glutamic acid. The participation of heme in the metabolic pathways of phycobilin has been suggested as well, due the structural similarities between heme and bilins. By inhibiting an enzyme involved in the heme metabolic pathway [11-14], production of phycocyanin was stopped, suggesting that heme was a direct precursor of phycobilins. The metabolism of phycobilins was deduced to be similar to the metabolism of heme. However, it includes the cleavage of the heme macrocyle by a heme oxygenase. The resulting open tetrapyrrole chain: biliveridin, BV IXα, is either bound to a phycobiliprotein or further reduced by ferredoxin-dependent bilin reductases leading to various phycobilin structures.

Attachment of bilins to biliproteins is a post-translational process that leads to the native structure of phycobiliproteins. The phycobilins are covalently attached to phycobiliprotein through a thioether bond. The site of attachment in phycobiliproteins is highly conserved. In the C-phycoerythrin alpha subunit, there are two conserved sites of attachment: Cys^{82} and Cys^{139}. There are three conserved sites of attachment in the C-phycoerythrin beta subunit, Cys^{80} and Cys^{165} as well as double attachment at Cys^{48} and Cys^{59}.

The main mechanism involved in phycobilin attachment involves enzyme catalysis. The first such lyase to be characterized was from S*ynechococcus* PCC 7002. It is a heterodimeric lyase composed of CpcE and CpcF proteins that catalyzed the attachment of PCB to Cys-α84 of phycocyanin [15]. It was also shown to catalyze both attachment and release of PCB. It is further capable of ligating PEB to the C phycocyanin alpha subunit, but with a reduced efficiency [15]. Another type of bilin lyase enzyme has been recently reported; this heterodimer is composed of CpcS and CpcU subunits [16,17], and it catalyzes the attachment of PCB to Cys-α84 in nearly all apo-phycobiliproteins [18]. A related protein called CpeS was shown to ligate PCB to Cys-84 of the beta subunit of C-phycocyanin [19]. A third family of bilin lyases called T-type lyase has been investigated and it was found that it catalyzes the specific addition of PCB at Cys-153 of CpcB from *Synechococcus* PCC 7002 [20]. *In vitro,* spontaneous attachment of PEB and PCB has been reported as well [21,22].

Phycobilins are attached exclusively through their A-ring [23-25]. Attachment to the D-ring does not occur, although it was reported [26,27] but was later retracted [24]. Double attachment at both A- and D-rings has been reported [21, 25, 26] in phycoerythrins. Structural studies of peptide-chromophore complexes determined that the carbon 3^1 was involved in the carbon-sulfur covalent thioether bond in A-ring attachment, whereas carbon 18^1 was implicated in D-ring attachment. The configuration of the resulting asymmetric carbon was determined using X-ray Crystallography [3].

This project is a collaboration work between the Department of Biological Sciences and the New Orleans Center for Mass Spectrometry Research. Dr. Wendy Schluchter's research group has been heavily involved in investigating and characterizing new bilin lyases. Our role as collaborators was to develop analytical methods to help identify sites of attachment of phycobilins on phycobiliprotein subunits. Mass spectrometry is a technique of choice used to identify and localize post-translational modifications (PTMs). The

development of soft ionization techniques, such as MALDI and ESI, and the consistent improvement of analyzer resolution, detector sensitivity and tandem mass spectrometry capabilities made the identification and localization of PTMs possible [28]. We developed an off-line LC-MALDI TOF/TOF method that enabled us to investigate the site-specific attachment of PEB to C-phycoerythrin alpha and beta subunits. This attachment was catalyzed by two newly reported enzymes: CpeY/Z and CpeS. The enzyme complex CpeY/Z shows some similarities in its amino acid sequence to CpcE/F enzyme complex, whereas CpeS shows some similarities in its amino acid sequence to the CpcS/U lyase complex. Further investigation of bilin lyase activity will enable us to conclude if the sequence similarities of these enzymes resulted in similar bilin activities.

3.3 Materials and Methods

Construction of expression vectors, heterologuous expression, purification of recombinant proteins, protein digestion as well as HPLC separation were all protocols performed by Dr. Avijit Biswas.

Construction of expression vectors, heterologuous expression and purification of recombinant proteins. Detailed protocols are available in Biswas et al, 2011 [29]**.** All expression constructs newly produced for this study were sequenced at the W. M. Keck Conservation and Molecular Genetics Laboratory (University of New Orleans) to confirm that no mutations had been introduced during PCR amplification and cloning. Each gene was amplified by PCR from *F. diplosiphon* chromosomal DNA, and each resulting amplicon was cloned into a Duet vector after digestion with restriction enzymes.

Heterologous expression and purification of recombinant protein. Expression plasmids were co-transformed into *E. coli* cells as required, and colonies were selected on Luria-Bertani (LB) plates. To produce PEB using the pPebS expression plasmid, a 50-ml overnight starter culture was added to 1 L of LB medium with the appropriate combination of antibiotics. This culture was shaken at 37 °C for 4 h until the optical density reached $OD_{600\ nm}$ = 0.6. Cells were incubated with shaking at 190 rpm at 18 °C for another 16 h before they were harvested by centrifugation at 10,000 × *g* for 10 min. Cell pellets were stored at -20 °C. *E. coli* cells containing recombinant proteins were thawed and resuspended. The cells were lysed, and the hexa-histidine-tagged recombinant proteins were purified as previously

described [30]. The recombinant protein(s) were exhaustively dialyzed overnight at 4 °C to remove the imidazole introduced during elution.

Tryptic Digestion of Phycoerythrin. CpeA-PEB and CpeB-PEB retrieved from the nickel-nitrilotriacetic acid-superflow-affinity column were subjected to tryptic digestion following a protocol previously published [31]. The purified holoprotein were exhaustively dialyzed against 2mM sodium phosphate buffer, pH 7.0, 1 mM 2-mercaptoethanol and then concentrated by ultrafiltration through an Amicon YM10 (Millipore, Billercia, MA). Concentrated samples were diluted 1:3 and titrated to pH 2.0 with 1N HCl. The solution was incubated for 45min in dark at room temperature for complete unfolding of protein. Trypsin was added from a 5 mg/ml stock solution in 1 mM HCl. Ammonium bicarbonate was added in 0.1 M and the mixture was titrated to pH 7.5 with 1N NaOH. The digested mixtures were incubated at 30 °C for 2h in dark. An additional aliquot of trypsin was added and incubated for another 2h. The reaction was quenched by adding glacial acetic acid. The mixtures were passed through a C-18 sep-pack cartridge. The eluted sample was vacuum-dried and stored at -20C for HPLC separation.

High Performance Liquid Chromatography. Tryptic peptides were separated on a C18 reverse-phase HPLC column (5µm X 10mm X 250 mm; Water Corp., Milford, MA) using a Waters HPLC equipped with a 600E pump and a photodiode array detector. The peptide separation was carried out as described by Arciero et al [31] using 0.1 M Na-phosphate, pH 2.1 as (Solvent A) and acetonitrile (Solvent B). The bilin peptides were eluted with an increasing concentration of acetonitrile and were monitored at 560 nm. The eluted sample was vacuum-dried and kept at -20 C for mass spectrometry analysis.

MALDI MS and MS/MS. MALDI MS and tandem MS experiments were performed on an Applied Biosystems (Foster City, CA) / MDS Sciex (Concord, Ontario) 4800 MALDI ToF/ToF. Mass spectral acquisitions were obtained in the reflectron mode using a Nd:YAG laser operated at 355 nm. The matrix used was α-cyanohydroxycinnamic acid (Sigma Aldrich) at 15 mg/mL in 50 %, v/v, acetonitrile (Sigma Aldrich) / 0.1%, v/v, trifluoroacetic acid (Sigma Aldrich). A volume of 2 µL from each fraction was mixed with 2 µL of matrix. The mixture was homogenized and then 0.75 µL was spotted on a MALDI 384-well plate and air dried.

3.4 Result and Discussion

The CpeA-PEB was submitted to tryptic digestion, followed by LC separation monitored at 550 nm (figure 3.1).

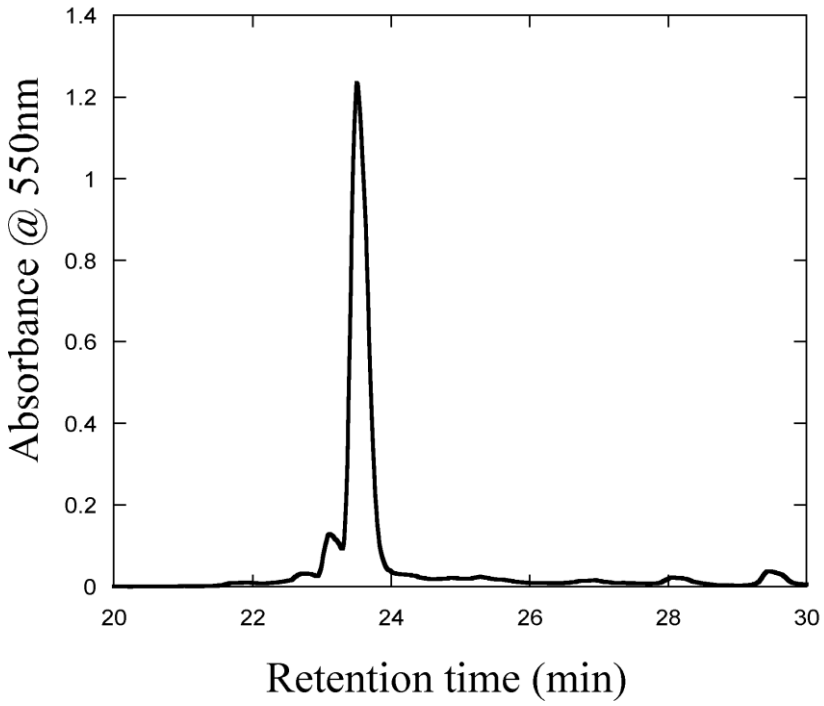


Figure 3.1. Chromatogram showing 550 nm absorbance vs retention time for the LC separation of a tryptic digest of C-phycoerythrin α subunit (CpeA). The choice of the absorption wavelength was based on the strong absorption of PEB-chromophores at 550nm. Two fractions were collected at t=23 min and t = 23.5 min.

The resulting fractions were submitted to MALDI MS (figure 3.2) and tandem MS analysis to identify unambiguously the location of the covalently attached PEB on HT (Histidine Tagged) CpeA. Tandem mass spectrometry using MALDI MS of peptides resulting from the tryptic digestion of the covalent complex HT-CpeA-PEB was performed to accomplish this.

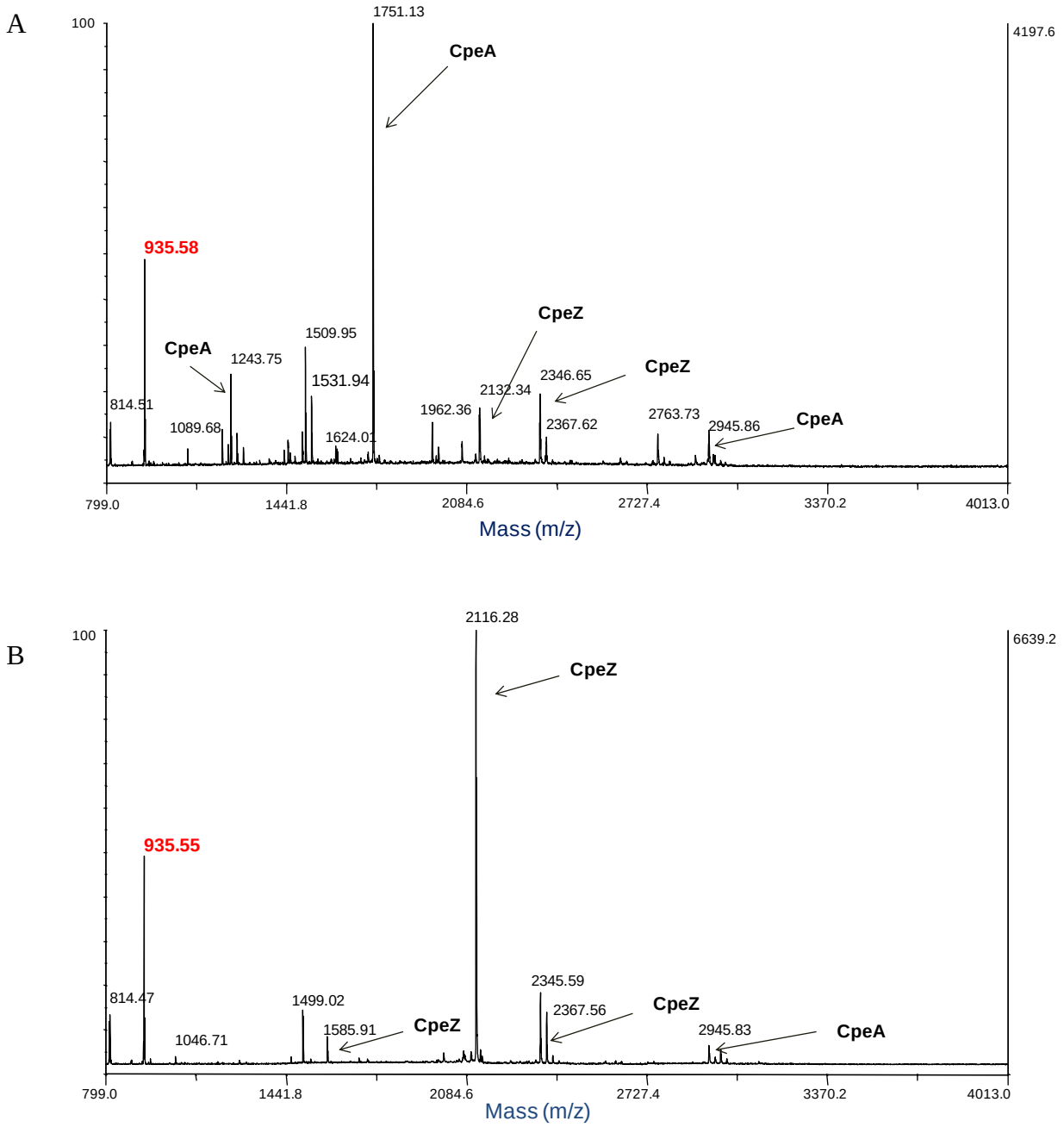


Figure 3.2. MALDI MS spectra of HPLC fractions at t= (A) 23 min, (B) 23.5 min. Arrows point to peptides identified using MASCOT database searching.

We sought to identify one or more peptides produced upon digestion that contained ligated PEB, notably a peak at m/*z* 935 appeared in all fractions. Figure 3.3.A shows the MS/MS spectrum of the precursor at m/*z* 935.

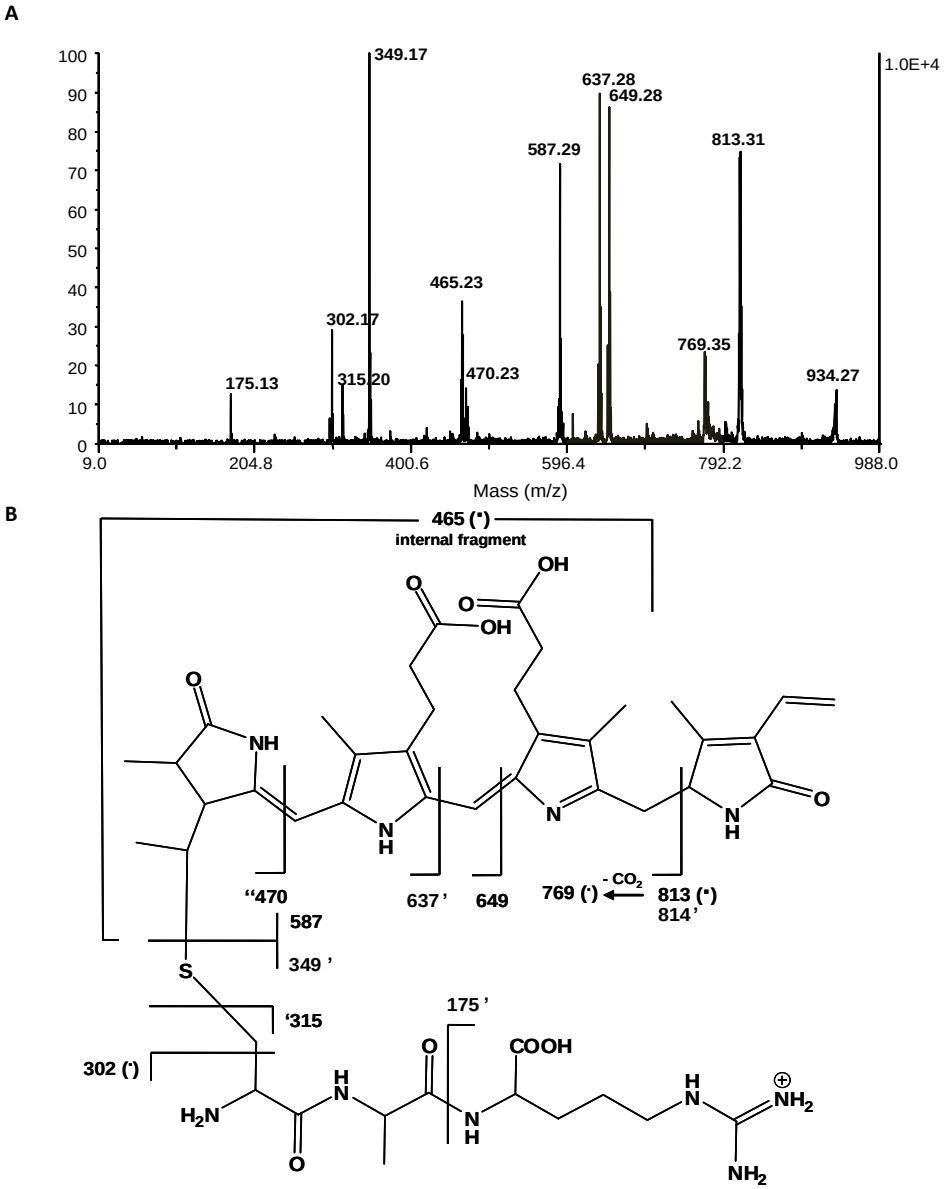


Figure 3.3. Mass spectrometric analyses of tryptic peptides of CpeA-PEB produced with CpeY and CpeZ. (A) MALDI MS/MS spectrum of the m/*z* 935 precursor ion derived from peptides resulting from the tryptic digestion of the covalent complex CpeA-PEB. This m/*z* 935 precursor ion was deduced to be a peptide fragment with a covalently bound PEB chromophore. (B) Fragmentation pattern and corresponding mass assignments for data in panel A. A tick mark prior to number, e.g., '470, indicates one hydrogen has been transferred to the departing neutral upon cleavage. A tick mark after number, e.g., 814', indicates the transfer of one hydrogen to the formed ion. A (·) indicates a radical ion [35].

In particular, there are two main peaks of interest at m/z 587 and 349. The peak at m/z 587 is attributed to protonated PEB. The peak at m/z 349 matches with a tripeptide, whose sequence is (K) CAR (D) and which contains Cys^{82}. The remaining major peaks were also assignable in a manner consistent with covalent attachment of PEB to Cys^{82}. The scheme in Figure 3B summarizes the structures of the assigned peaks. The above spectral interpretation suggests that, by applying sufficient collision energy in tandem mass spectrometry experiments, it was possible to break the thioether bond and separately detect the chromophore (m/z 587) and the peptide (m/z 349). The mechanism of fragmentation leading to the detection of the protonated chromophore and peptide is described in figure 3.4. The structure of the chromophore, which is highly conjugated, favored the formation of product ions that were stabilized by resonance. The large number of peaks enabled a thorough structural elucidation of the peptide-PEB covalent complex.

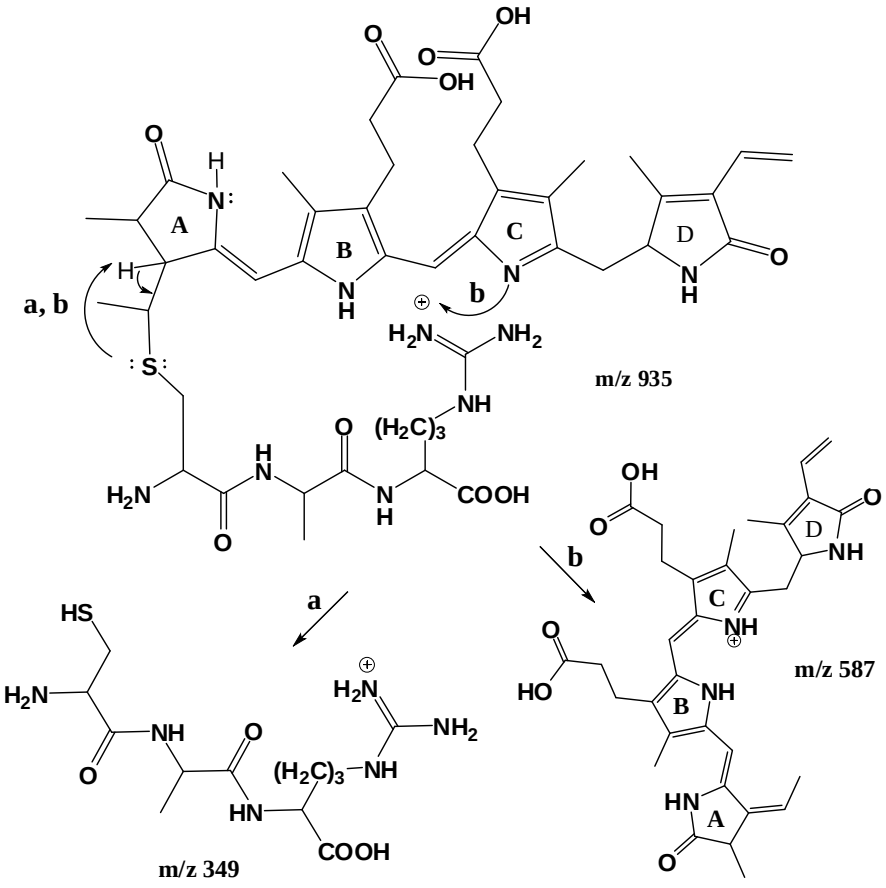
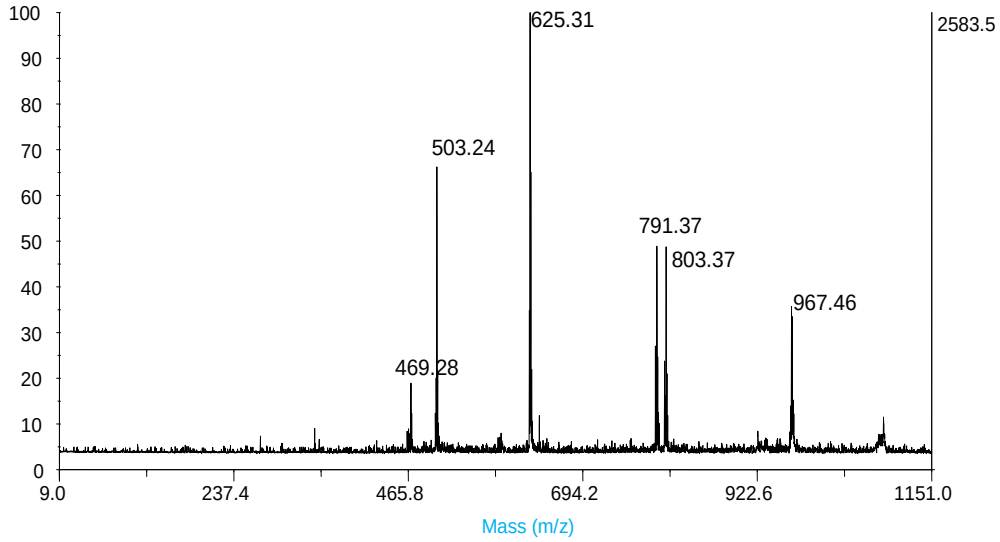


Figure 3.4. Decomposition pathways of the protonated PEB-tripeptide complex leading to formation of protonated tripeptide (m/z 349, pathway a) and protonated PEB (m/z 587, pathway b).

In a follow-up analysis, we specifically investigated whether attachment of PEB occurred at Cys^{139}. We were able to identify a peptide at m/z = 503 that contained the unmodified cysteine at position 139 (sequence: (R) GCAPR (D)). If attachment of PEB occurred to Cys^{139}, then it should have been possible to detect a chromopeptide with m/z = 1089 (503 + 586). An ion with this m/z ratio was indeed detected in one fraction, the product ion mass spectrum from this m/z 1089 (Figure. 3.5.A) precursor yielded a minor peak at m/z 587 and a peak at m/z 503 corresponding to the neutral loss of 586. This latter peak thus represents (R) GC^{139}APR (D) that has lost the bilin moiety. Furthermore, a comparison of the peak heights observed in MALDI mass spectra that simultaneously contain m/z 1089 and 935, revealed that the former peak was quite minor compared to the latter. Although peak heights observed in MALDI mass spectra are not strictly reliable for quantification, it appears that Cys^{82} was the highly favored site of attachment.

Attachment of PEB to CpeA catalyzed by CpeY/Z was also investigated by Site Directed Mutagenesis by our collaborators. Cysteines at postion 82, 139 were replaced by a serine. The spectral features of three different mutants: C82S, C139S and C82S/C139S were investigated. Absorption and fluorescence measurement on the mutants showed that only one mutant, the C139S, had spectral features similar to the HT-CpeA. This result proves that CpeY/Z attaches PEB to the mutant that had the cysteine 82 unchanged. This result clearly shows the site specificity of CpeY/Z to Cys^{82} of CpeA. Mass spectrometric results confirmed the Site Directed Mutagenesis results by identifying a chromophorylated peptide, at m/z 935, containing Cys^{82}, and the detection of another "minor" chromophorylated peptide at m/z 1089 containing Cys^{139} could be due to autocatalytic attachment (non-enzymatically attached). The results of both experiments showed that, CpeY/Z primarily ligates PEB to Cys^{82}.

A



B

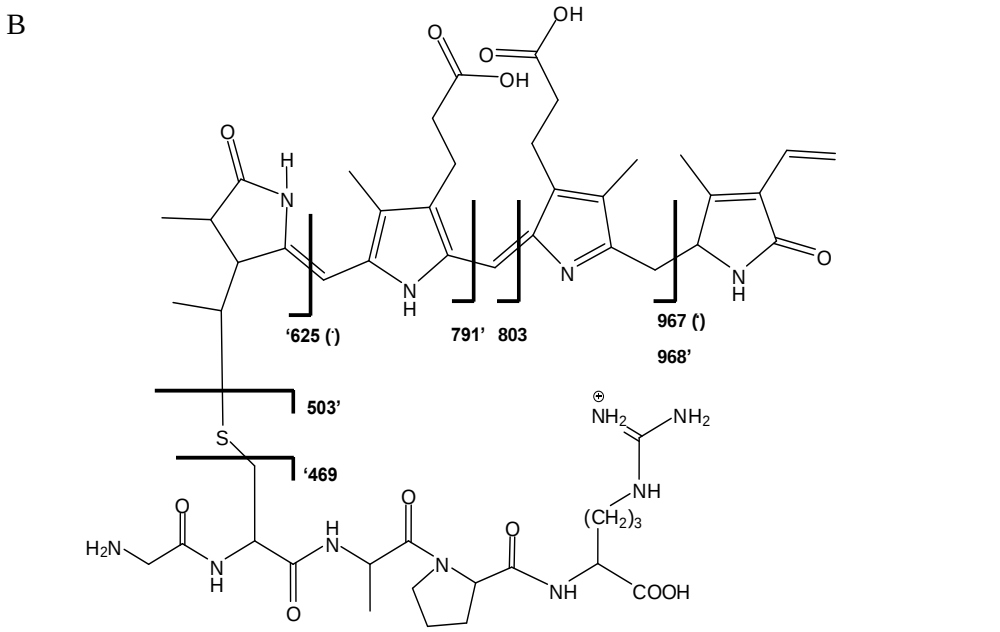


Figure 3.5. (A) MALDI MS/MS spectrum of the precursor ion at m/z 1089, which was deduced to be a peptide fragment with a covalently bound PEB chromophore. This peptide holding PEB was derived from trypsin digestion of the HT-CpeA-PEB produced in the presence of CpeY and CpeZ. The MS/MS spectrum contains a peak of interest at m/z 503. The peak, resulting from a neutral loss of 586, was attributed to a peptide containing a cysteine at position 139. The sequence of the peptide is (R) GCAPR (D). The peak corresponding to protonated PEB, which is detected at m/z 587, was not detected in the spectrum shown in this figure. However, when applying a higher acceleration voltage the peak is visible. (B) Peak assignments of product ion spectrum corresponding to the precursor protonated PEB-peptide (derived from CpeA) complex [36].

Recombinant HT-CpeB (non-variant) produced with CpeS and the enzymes for PEB synthesis was purified and then subjected to trypsin digestion. The resulting tryptic peptides were separated by HPLC chromatography on a reversed-phase C18 column. In the chromatogram monitored at 550 nm to detect peptides with bound PEB, two peaks, eluting at 23 and 24 min, were observed and collected. MALDI MS and tandem MS was used to identify the peptides from these two peaks. Figure 3.6.A shows the MS/MS spectrum of the precursor at m/z 1250. The peaks at m/z 587 and 664 were the most informative. The m/z 587 peak corresponds to protonated PEB as previously discussed. The peak at m/z 664 matched a peptide containing a cysteine at position 80, (R) MAAC^{80}LR (D). The scheme in figure 3.6.B summarizes the structures of the assigned peaks. A review of the tandem mass spectra did not show attachment to any other peptide. These results confirm that the CpeS bilin lyase specifically attaches PEB to Cys^{80} of CpeB and to no other Cys residues.

Our mass spectrometry results build upon previous studies investigating the location of the attachment site of phycobilins [26, 27]. The PEB-peptide resulting from tryptic digestion of the β-PE was detected at m/z 1250 [18, 32]. Fragments resulting from tandem mass spectrometric experiments such as protonated free PEB (m/z 587) [25] and the tri-pyrrole fragment that results from the loss of pyrrole ring D (m/z 464) were also reported [33]. Using an off-line MALDI ToF/ToF method, Wiethaus and coworkers [34] were very recently able to locate the site of ligation of PEB on the β-PE subunit from *P. marinus* based upon a major peak corresponding to PEB loss from a sequenced tryptic peptide containing Cys^{82}.

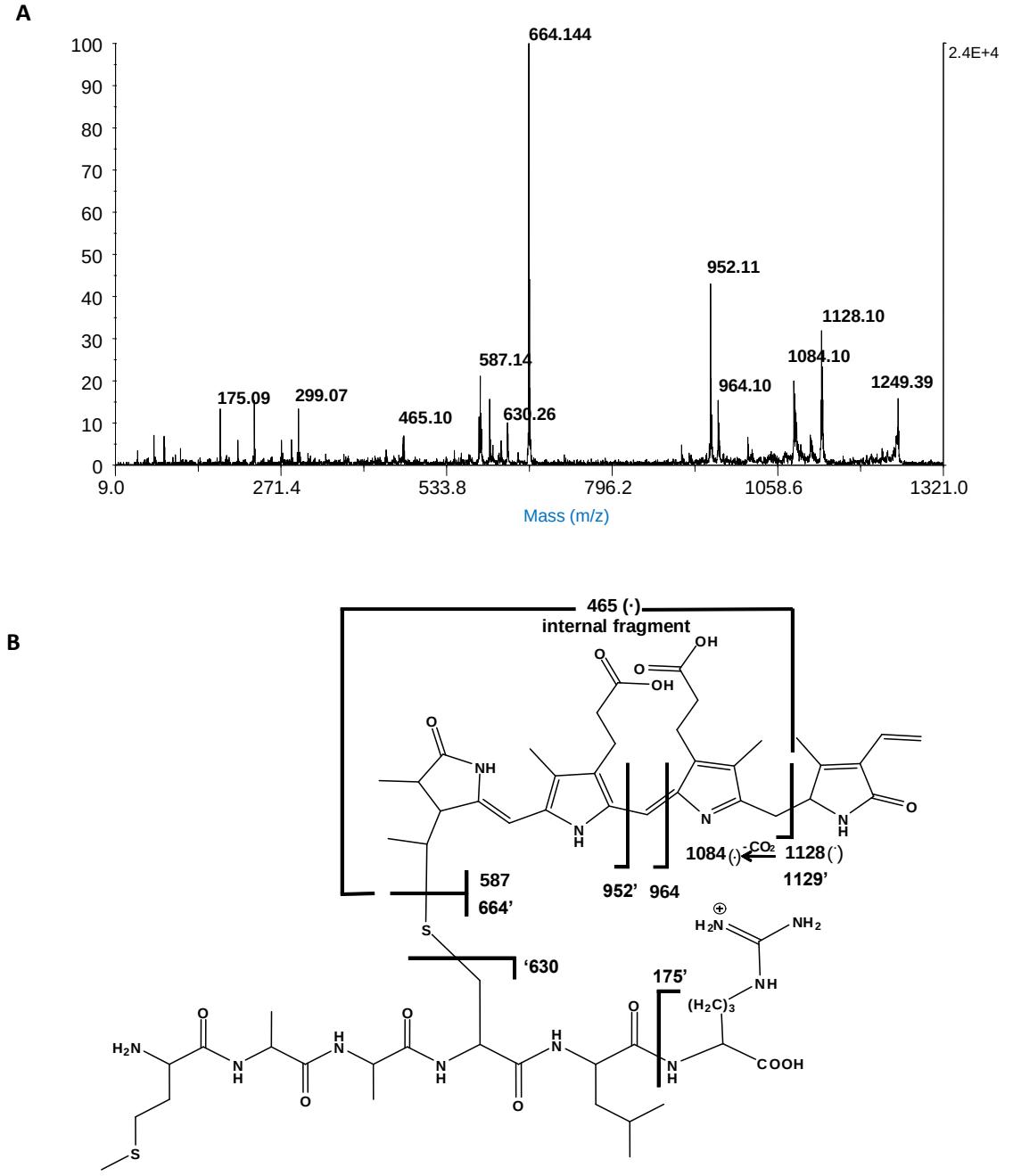


Figure 3.6. (A) MALDI MS/MS spectrum of the precursor ion at m/z 1250, deduced to be a peptide fragment with a covalently bound PEB chromophore, and which was derived from trypsin digestion of the HT-CpeB-PEB produced in the presence of CpeS. The MS/MS spectrum contains two peaks of interest at m/z 664 and m/z 587. The peak at m/z 664 was attributed to a peptide containing a cysteine at position 80. The sequence of the peptide is (R) MAACLR (D). The second peak at m/z 587 was attributed to protonated PEB. (B) Peak assignments of product ion spectrum corresponding to the precursor protonated PEB-peptide (derived from CpeB) complex.

The molecular structure of the tetrapyrrole, PEB, has an extended "π-conjugated" system, and may exist as three different tautomers. These isomeric structures, which exist in a dynamic equilibrium, differ only in that the pyrrole rings carrying the imino and amino nitrogens have changed. Figures 3.3, 3.4, 3.5 and 3.6 show structures of PEB and PEB attached to tryptic peptides derived from CpeA and CpeB. The tautomeric forms shown are based upon previously published work [26, 27, 32] that relied upon NMR to assign the predominant tautomer.

3.5 Conclusions

A newly developed off-line LC-MALDI MS and tandem MS method enabled us to investigate the site-specific attachment of phycobilins catalyzed by various bilin lyase enzymes to C-phycoerythrin alpha and beta subunit. We successfully contributed to the full characterization of a new set of bilin lyase, CpeY/Z and CpeS, by identifying their specific site of ligation of PEB to CpeA and CpeB, respectively. CpeY/Z catalyzes the specific attachment of PEB to Cys^{82} of the C-Phycoerythrin alpha subunit. Results of both mass spectrometry and site directed mutagenesis show that CpeY/Z primarily attaches PEB to Cys^{82}. CpeS attaches PEB exclusively to Cys^{80} of the C-phycoerythrin beta subunit. CpeS has been introduced as a nearly universal lyase for Cys-82-equivalent sites in cyanobacteria. However, comparison of the attachment efficiency of PEB to α-Cys^{82} when catalyzed by CpeY/Z and CpeS strongly indicates that CpeY/Z is the principal bilin lyase responsible for the attachment to α-Cys^{82} and not CpeS as was previously published.

3.6 References

[1] Glazer, A. N., Light Guides: directional energy transfer in a photosynthetic antenna. *The Journal of Biological Chemistry* **1989**.264, 1-4

[2] Scheer, H.; Zhao, K. –H. Biliprotein maturation: the chromophore attachment. *Molecular Microbiology* **2008,** 68, 263-276

[3] Scheer, H., Biliproteins. *Angewandt Chemie International Edition in English.***1981**, 20, 241-261

[4] Shirmer, T.;Bode, W.; Huber, R., Refined three-dimensional structures of two cyanobacterial C-phycocyanins at 2.1 and 2.5A resolution: A common principle of phycobilin-protein interaction. *Journal of Molecular Biology* **1987**, 196, 677-695

[5] Duerring, M.; Schmidt, G. B.; Huber,R.; Isolation, crystallization, crystal structure analysis and refinement of constitutive C-phycocyanin from the chromatically adapting cyanobacterium *Fremyella diplosiphon* at 1.66A resolution. *Journal of Molecular Biology* **1991**, 217, 577-592

[6] Crespi, H. L.; Boucher, L. J.; Norman, G. D.; Katz, J.J., The structure of phycocyanobilin. *The Journal of the. American. Chemical. Society* **1967**, 89, 3642-3643

[7] Cole, W. J; Chapman, H. W.; Siegelman, H. W., The structure of phycocyanobilin. *Journal of the American Chemical Society* **1967**, 89, 3643-3645

[8] Chapman, D.J; Cole, W. J.; Siegelman, H. W., The structure of phycoerythrobilin. *Journal of the American Chemical Society* **1967**, 89, 5976-5977

[9] Frankenberg-Dinkel, N.; Terry, M. J.; Synthesis and Role of Bilins in Photosynthetic Organisms. In Warren, M. J.; Smith, A. G. Tetrapyrroles: Molecular Biology Intelligence Unit (pp 208-220). **2009**, Springer, NY publisher

[10] Troxler, R.; Synthesis of bile pigments in plants. Formation of carbon monoxide and phycocyanobilin in wild type and mutant strains of the alga, Cyanidium caldarium. *Biochemistry* **1972**, 11, 4235

[11] Beale, S. I.; Chen, N. C.; *N*-Methyl Mesoporphyrin IX Inhibits Phycocyanin, but not Chlorophyll Synthesis in *Cyanidium caldarium. Plant Physiology* **1983,** 71, 263-268

[12] Beale, S. I.; Biosynthesis of Phycobilins. *Chemical Reviews* **1983**, 93, 785-802

[13] Beale, S. I.; Enzymes of chlorophyll biosynthesis. *Photosynthesis Research* **1999**, 60, 43-73

[14] Brown, S. B.; Holroyd, J. A.;Troxler, R.F.; Offner, G. D. Bile pigment synthesis in plant. Incorpotation of haem into phycocyanobilin and phycobiliproteins.in Cyanidium caldarium. *Biochemical Journal* **1981**, 194,137-147

[15] Zhou, J.; Gasparich, G.E.; Stirewalt, V.L; De lorimier, R.; Bryant, D.A. The cpcE and cpcF genes of *synechococcus sp.* PCC 7002. Construction and phenotypic characterization of interposon mutants. *Journal of Biological Chemistry*. **1992**. 267, 16138-16145

[16] Shen, G.; Saunee, N.A.; Gallo, E; Begovic, Z.; Schluchter, W.M.; Bryant, d.A.; Identification of novel phycobiliprotein lyases in cyanobacteria. In *PS 2004 Light-Harvesting*

systems Workshop. Saint Adele, quebec, Canada: International Society of Photosynthesis Research, pp. 14-15

[17] Saunee, N. A., Williams, S. R., Bryant, D. A., Schluchter, W. M. Biogenesis of phycobiliproteins. *II. CpcS-I and CpcU comprise the heterodimeric bilin lyase that attaches phycocyanobilin to cys-82 of β-phycocyaninand cys-81 of allophycocyanin subunits in Synechococcus sp. PCC 7002*. *The Journal of Biological Chemistry*. **2008**, 283, 7513-7522

[18] Zhao, K. H; Su , P.; Tu, J.M; Wang, X.; Liu, H., PLoscher, M.; Eichacker, L.; Yang, B., Zhou, M.; Scheer, H.; Phycobilin:cystein-84 biliprotein lyase, a near universal lyase for cystein-84-binding sites in cyanobacterial phycobiliptoein. *Proceedings of the National Academy of Science of the United States of America.* **2007**, 104, 14300-14305

[19] Zhao, K. H., Su, P., Li, J., Tu, J. M., Zhou, M., Bubenzer, C., Scheer, H. Chromophore attachment to phycobiliprotein beta-subunits: phycocyanobilin: cystein-beta84 phycobiliprotein lyase activity of CpeS-like protein from Anabaena Sp. PCC7120. *The Journal of Biological Chemistry*. **2006**, 281, 8573-8581

[20] Shen, G.; saunee, N.A.; Williams, S. R.; Gallo, E.F.; Schluchter, W.M; Bryant, D.A.; Identification and characterization of a new type of new class of bilin lyase. The cpcT gene encodes a bilin lyase responsible for attachment of phycyanobilin to cys-153 on the beta subunit of phycocyanin in *Synechococcus sp.* PCC 7002. *Journal of Biological Chemistry*. **2006.** 281, 17768-17778

[21] Schluchter, W. M.;Glazer, A.N.; Biosynthesis of phycobiliproteins in cyanobacteria. In *The Phototrophic Prokaryotes*. Peschek, G.A.; Loffelhardt, W.; Schmetterer, G. (editors).New York: Kluwer/Plenum Press, pp. 83-95

[22] Shcluchter, W.M; Bryant, D. A., Analysis and reconstitution of phycobiliproteins. Methods for the characterization of bilin attachment reactions. In heme, Chlorophyll and Bilins. Smith, A.G; Witty, M., (editors). Totowa, NJ: Humana Press, pp. 311-334

[23] Schoenleber, R. W.; Lundell, D. J.; Glazer, A. N.; Rapoport, H., Bilin Attachment sites in the α and β Subunits of B-phycoerythrin. Structural studies on the singly linked Phycoerythrobilins. *The Journal of biological Chemistry*. **1984**, 259,5485-5489

[24] Schoenleber, R. W.; Leung, S. L.; Lundell, D. J.; Glazer, A. N.; Rapoport, H., Chromopeptides from Phycoerythrins. Structure and Linkage of Phycoerythrobilin Tryptic Tripeptide Derived from a B-Phycoerythrin. *The Journal of the American Chemical Society.***1983**, 105, 4072-4076

[25] Lagarias, J. C.; Klotz, A. V.; Dallas, J.; Glazer, A. N.; Bishop, J. E.;O'connell, J. F.; Rapoport, H., Exclusive A-ring Linkage for Singly Attached Phycocyanobilins and Phycoerythrobilins in Phycobiliproteins. Absence of singly D-ring-linked Bilin. *The Journal of Biological Chemistry.***1988,** 263, 12977-12985

[26] Bishop, J. E.; Lagarias, J. C.; Nagys, J. O.; Schoenelber, R. W.; Rapoport, H.; Klotz, A. V.; Glazer, A. N., Phycobiliprotein-Bilin Linkage Diversity.Structural study on A- and D-ring-linked Phycocyanobilins. *The Journal of Biological Chemistry*. **1986**, 261, 6790-6796

[27] Klotz, A. V.; Glazer, A. N.; Bishop, J. E.; Nagys, J. O.; Rapoport, H.; Phycobiliprotein-Bilin Linkage Diversity.Structural study on A- and D-ring-linked Phycerythrobilin. *The Journal of Biological Chemistry*. **1986**, 261, 6797-6805

[28] Mann, M.; Jensen, O. N.; Proteomic analysis of post-translational modifications. *Nature Biotechnology*. **2003**, 21, 255-261

[29] Biswas, A., Boutaghou, M. N., Alvey, R. M., Kronfel, C. M., Cole, R. B., Bryant, D. A., Schluchter, W. M. Characterization of the Activities of the Cpe Y, Cpe Z and CpeS Bilin Lyases in Phycoerythrin Biosynthesis in *Fremyella diplosiphon* Strain UTEX 481. *The Journal of Biological Chemistry*. **2011**, 286, 35509-35521

[30] Shen, G., Saunee, N. A., Shervonda, R. W., Gallo, E. F., Schluchter, W. M., Bryant, D. A. Identification and Characterization of a new Class of Bilin Lyase. *The CpcT gene encodes a bilin lyase responsible for attachment of phycocyanobilin to Cys-153 of the β-subunit of phycocyanin in Synechoccocuss SP. PCC 7002*. *The journal of Biological Chemistry.* **2006**, 281, 17768-17778

[31] Arciero, D. M.; Bryant, D. A.; Glazer, A. N.; *In Vitro* Attachment of Bilins to Apophycocyanins. Specific covalent adduct formation at cysteinyl residues involved in phycocyanobilin binding in C-phycocyanin. *The Journal of Biological Chemistry*. **1988**, 263, 18343-18349

[32] Isailovic, D., Sultana, I., Phillips, G. J., and Yeung, E. S. Isolation and characterization of R-phycoerythrin subunits and enzymatic digests. *Journal of Chromatography A.* **2004**, 1051,119-130

[33] Bishop, J. E., Rapoport, H., Klotz, A. V., Chan, C. F., Glazer, A. N., Guglistaller, P., and Zuber, H. Chromopeptide from Phycoerythrocyanin. Structure of the three bilin groups. *Journal of the American Chemical Society.* **1987**, 109,875-881

[34] Wiethaus, J., Busch, A. W. U., Kock, K., Leichert, L. I., Herrmann, C., and Frankenberg-Dinkel, N. CpeS Is a Lyase Specific for Attachment of 3Z-PEB to Cys-82of β-phycoerythrin from *Prochlorococcus marinus* MED4, *The Journal of Biological Chemistry.* **2010,** 285, 37561-37569

[35] Boutaghou, M. N., Biswas, A., Cole, R. B., Schluchter, W. M. "Localization of the Specific Site of Enzyme-Mediated Covalent Binding of Phycoerythrobilin (PEB) to C-Phycoerythrin Alpha Subunit." Proceedings of the 58th ASMS Conference on Mass Spectrometry and Allied Topics, Salt Lake City, UT, May 26, 2010. (publisher: ASMS)

[36] Boutaghou, M. N., Biswas, A., Schluchter, W. M., Cole, R. B. "Localization of the Specific Site of CpeS Enzyme-Mediated Covalent Binding of Phycoerythrobilin (PEB) to the C-Phycoerythrin Beta Subunit." Proceedings of the 59th ASMS Conference on Mass Spectrometry and Allied Topics, Denver, CO, June 9, 2011. (publisher: ASMS)

Chapter 4

Investigating A-ring versus A- and D-ring attachment of phycoerythrobilin (PEB) chromophores using MALDI mass spectrometry

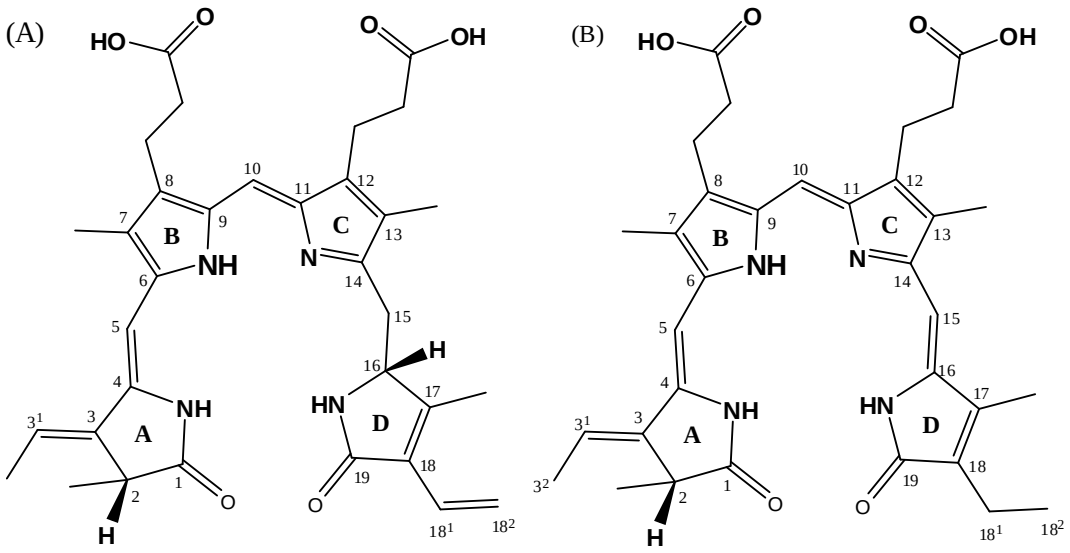
4.1 Abstract

Bilin chromophore attachment to phycobiliproteins. is an enzyme-catalyzed post-translational modification reaction. Bilin-lyases attach bilin chromophore to their host proteins through a thioether bond between the chromophore and a cysteine moiety. Bilin chromophores are attached to the proteins through the carbon 3^1 of the bilin. Double attachment occurs as well and in this case, carbons 3^1 and 18^1 of the bilin are both forming covalent linkages. There is a mass spectrometric limitation when examining tryptic peptides containing two (or more) cysteines if one seeks to ascertain whether chromopeptides are singly or doubly attached. The problem is that singly and double attached chromopeptides appear at the same m/z value, thus, up until the present, only NMR analysis has been used to determine whether the chromophore is singly or doubly attached. We report in this work a new fast and accurate method for discriminating singly from doubly attached chromophores using MALDI TOF/TOF mass spectrometry. This method was deduced from spectral analysis of chromopeptides resulting from *in-vitro* and *in-vivo* attachment of bilin chromophores to phycobiliproteins. Distinction is based on a difference corresponding to a characteristic neutral loss that appears in the MALDI-TOF mass spectrum only when the bilin is singly attached.

4.2 Introduction

Light harvesting organisms, such as cyanobacteria, contain two types of photosystems: photosystem I and photosystem II. Phycobiliproteins are the major light absorbing pigment in photosystem II and they reorganize into a bigger complex, called a phycobilisome (PBS) [1,2]. Phycobiliprotein subunits will undergo post-translational modifications through the attachment of linear tetrapyrrolic molecules to highly conserved sites [3]. These tetrapyrrolic molecules are known as "phycobilins" or simply "bilins" and their attachment is catalyzed by bilin lyases. Neither free bilins, nor denatured phycobiliproteins are suitable photoreceptors since they absorb light inefficiently, and their excited states are very short-lived, leading to conversion of their excitation energy into heat. On the other hand, the phycobiliprotein subunit-bilin complex displays much improved photophysical properties: absorption coefficient of the chromophores is increased by one order of magnitude and the excited state lifetimes are prolonged by four orders of magnitude; this combination results in an excellent photoreceptor unit. The difference in properties appears to be due to the varying conformation of the bilin pigment. When the phycocyanobilin is free, it has a cyclohelical conformation that shows stronger near-UV band absorption and a weaker absorption in the visible [4-6].

Phycobilins such as PEB and PCB (Scheme 4.1) are attached to the phycobiliprotein exclusively through their A-ring [7-9]. D-ring attachment has not been shown to occur, although it was initially reported [10, 11] but was later retracted [9]. Double ring attachment has been reported [12-14] in phycoerythrins and involving ring A and D and two consecutive cysteines, i.e. the adjacent cysteine attached to the D-ring after initial A-ring attachment of the first cysteine. 1D and 2D NMR analysis was used in order to assess with certainty the double-attachment of the phycobilin-chromophore to the chromopeptide.

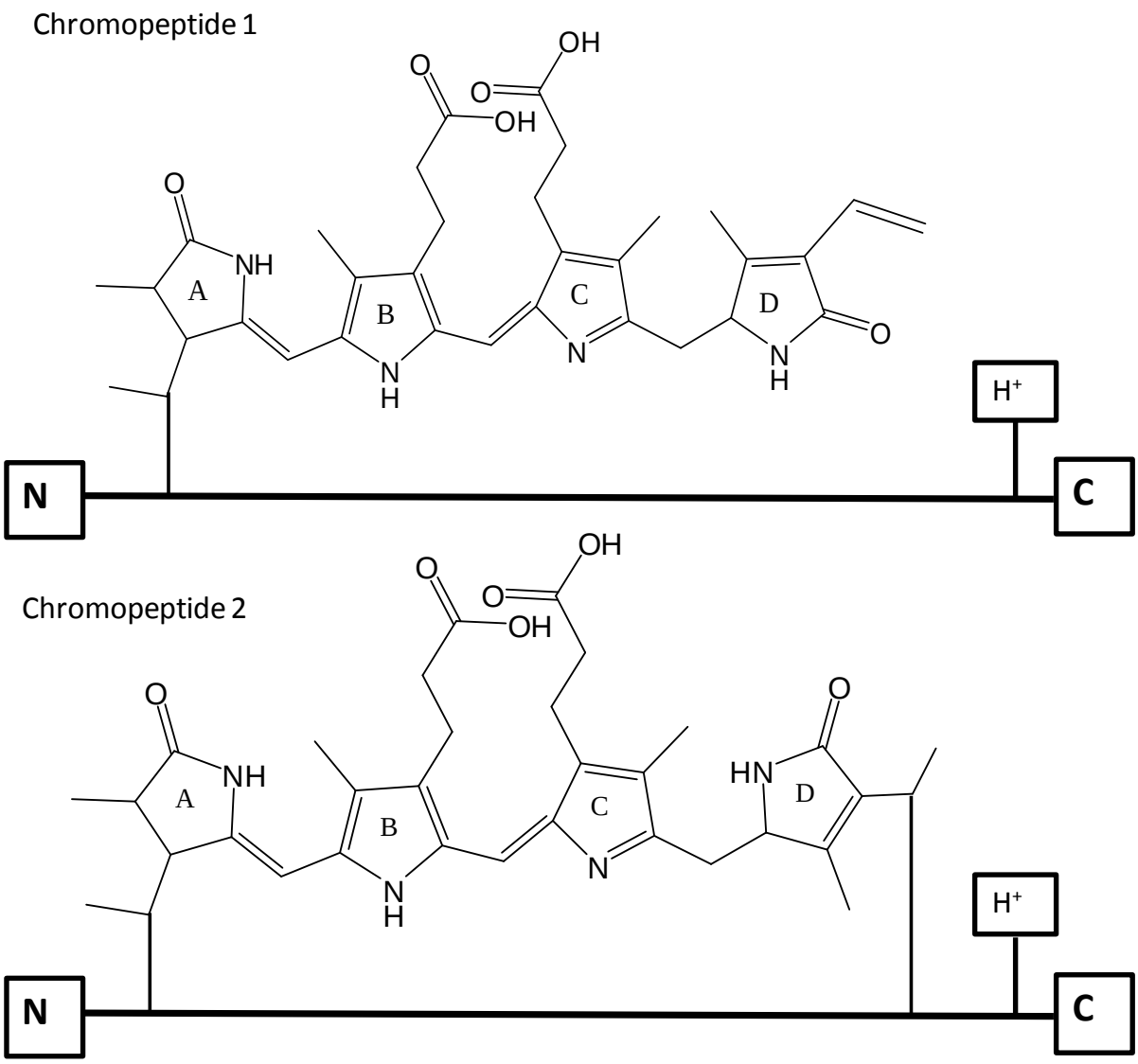


Scheme 4.1. (A) Structure of phycoerythrobilin (PEB) chromophore. (B) Structure of phycocyanobilin (PCB). Carbons and rings are numbered according to IUPAC recommendation. PEB chromophore is attached to C-phycoerythrin proteins (alpha and beta subunit) through a thioether bond between a cystein and the carbon 3^1. When double attachment is involved, the attachement occurs at carbon 3^1 and 18^1.

Mass spectrometry was used as well, but mainly to detect the mass of the PEB-chromopeptide complex. Mass spectrometry alone did not provide conclusive results to allow certainty in determining the number of site(s) of attachment. However, mass spectrometry remains a reliable analytical tool for determining accurately the exact location of enzyme-catalyzed attachment of phycobilins to phycobiliprotein. In particular, MALDI MS and tandem MS (coupled with off-line LC) have been successfully applied to the characterization of enzyme-catalyzed bilin attachment [15,16]. LC-ESI MS/MS was applied to the analysis of native phycobiliprotein, R-phycoerythrin and the mass of the whole protein was deduced based on an LC-(C4)-ESI mass spectrum deconvolution. Tryptic peptides were analyzed as well by LC (C18)-ESI-MS, which enabled the localization of the site of attachment of the PEB chromophore [17]. On the other hand, secondary ion MS was used for characterizing a chromopeptide containing a doubly attached PEB chromophore, through both carbons 3^1 and 18^1. The intact protonated chromopeptide complex was detected along with sodium and

potassium adducts [14]. A fragment at m/z 587 corresponding to protonated free phycoerythrobilin was observed as well. In terms of extracting information from chromopeptides containing an attached bilin, the main limitations of mass spectrometry are: 1) attachment at the A-ring or D-ring results in a product of the same m/z value so further information is needed to distinguish A-ring attachment from D-ring attachment; 2) the utility of tandem mass spectrometry for distinguishing A-ring attachment vs. D-ring attachment is impeded by the symmetry of the linear tetrapyrrole chromophore which may offer similar decompositions (neutral losses) from either end of the attached chromophore; and 3) singly attached chromophores have the same masses as doubly attached chromophores, hence, they cannot be distinguished based upon simple mass difference.

Under current investigation is the C-phycoerythrin beta subunit of the cyanobacterium *Fremyella diplosiphon*. On the subunit CpeB, phycoerythrobilin is known to be attached to Cys^{80} and Cys^{165} through the A-ring (carbon 3^1) and doubly attached to Cys^{48} and Cys^{59} through both rings A and D (carbon 18^1) [3]. In cases such as this, however, when investigating the double attachment of PEB to a single peptide containing two cysteines, MS measurements cannot confirm that PEB is attached to both cysteines (double attachment) as opposed to only one of them (single attachment) because both complexes have the same m/z value (Scheme 4.2). Moreover, the size of such complexes makes it impossible to obtain sufficient product ion yields to obtain good signals in collision induced decomposition (CID) MS/MS experiments; thus other techniques, e.g., NMR, must be used to ascertain the exact nature of the binding site.



Scheme 4.2. The mass of a PEB-peptide complex when PEB is doubly attached (through carbon 3^1 and carbon 18^1) or singly attached (through carbon 3^1) is the same. Therefore m/z (chromopeptide1) = m/z (chromopeptide 2)

We introduce in this work, a very simple and fast method for the investigation of the type of attachment (double or single) of one given phycobilin (phycoeryhrobilin, PEB) to Cys^{48} and Cys^{59} of the C-phycoerythrin beta subunit of *Fremyella diplosiphon*. The method was constructed based upon empirical observations made during MALDI MS analyses of PEB-attached tryptic peptides originating from in-vitro and in-vivo attachment of PEB.

4.3 Materials and Methods

Holo-protein extraction, digestion and HPLC separation were performed by Chritina Kronfel. HT-CpeA-PEB and HT-CpeB-PEB extraction, digestion and HPLC separation were performed by Dr. Avijit Biswas.

Holo-Phycoerythrin extraction. Holo-phycoerythrin alpha and beta subunits were extracted from *Fremyella diplosiophon* following a previously published protocol [18]. Cyanobacteria cells (50 g, wet weight) were suspended in 50 mL of 1M sodium acetate (pH 5.0) and were submitted to a cell disruptor (french press). The sonicates were pooled and centrifuged at 81,000 g. The supernatant was isolated and brought to 35% saturation with ammonium sulfate centrifuged at 16,000 g. The supernatant was cut, pellets were spun-down and resuspended in 100 mM sodium acetate and dialized against the same buffer overnight. The dialyzed samples were passed through a Sephadex G-100 column and fractions containing phycoerythrin were pooled and brought to 30% saturation with ammonium sulfate and centrifuged. Pellets were resuspended in 5 mM potassium phosphate (pH=7.0) and dialyzed overnight against the same buffer. The dialyzed samples were passed through a DEAE-Cellulose DE-52 (Whatman, GE healthcare, UK) column, pre-equilibrated with 5 mM potassium phosphate buffer. Fractions containing phycoerythrins were brought to 30% saturation with ammonium sulfate and centrifuged. The pellet containing purified phycoerythrin was resupended in a small volume of 100 mM sodium phosphate (pH 7.0) in 0.01% sodium azide and dialyzed against the same buffer. The dialyzed solution was concentrated and submitted to tryptic digestion.

Protein digestion. Purified CpeA-PEB, CpeB-PEB and holo-phycoerythrin were subjected to tryptic digestion following a previously published protocol [19]. The purified holoprotein was dialyzed against a 2 mM sodium phosphate buffer (pH 7.0) containing 1 mM 2-mercaptoethanol and then concentrated by ultrafiltration through an Amicon YM10 (Millipore, Billerica, MA) filter. Concentrated samples were diluted 1:3 and titrated to pH 2.0 with 1 N HCl. The solution was incubated for 45 min in the dark at room temperature for complete denaturation of protein. Trypsin was added from a 5 mg/mL stock solution in 1 mM HCl. 0.1 M ammonium bicarbonate was added, and the mixture was adjusted to pH 7.5 with 1 N NaOH. The digested mixtures were incubated at 30 °C for 2h in the dark. An additional aliquot of trypsin was added and incubated for another 2h. The reaction was quenched by adding 30% v/v glacial acetic acid. The mixtures were passed through a C-18 Sep-Pack (Waters, Milford, MA) cartridge. The eluted sample was vacuum-dried and stored at -20 °C prior to HPLC separation.

Liquid Chromatography. Tryptic peptides were separated on a C18 reversed-phase LC column (5µm X 10mm X 250 mm; Waters Corp., Milford, MA) using a Waters LC equipped with a 600E pump and a photodiode array detector. The peptide separation was carried out as described by Arciero et al. [19] using 0.1 M Na-phosphate buffer, pH 2.1 (Solvent A) and acetonitrile (Solvent B). The bilin peptides were eluted using a gradient of increasing concentration of acetonitrile (35 to 100%) and were monitored (absorbance) at 560 nm. The eluted samples were vacuum-dried and kept at -20 °C for mass spectrometry analysis.

MALDI MS and MS/MS. MALDI MS and MALDI MS/MS experiments were performed on an Applied Biosystems (Foster City, CA) / MDS Sciex (Concord, Ontario) 4800 MALDI ToF/ToF. Mass spectral acquisitions were obtained in the reflectron mode using a Nd:YAG

laser operated at 355 nm. The matrix used was α-cyanohydroxycinnamic acid (Sigma Aldrich, St. Louis, MO) at 15 mg/mL in 50%, v/v, acetonitrile (Sigma Aldrich) / 0.1%, v/v, trifluoroacetic acid (Sigma Aldrich). A volume of 2 µL from each fraction was mixed with 2 µL of matrix. The mixture was homogenized and then 0.75 µL was spotted on a MALDI 384-well plate and air dried.

4.4 Results and discussion

In a previous study, MALDI MS and MALDI MS/MS were successfully applied in combination with other techniques to deduce the site of attachment of PEB to C-phycoerythrin (alpha) subunits using newly developed bilin lyase enzymes [16]. Chromopeptides (peptides containing attached PEB) resulting from tryptic digestion of mature C-phycoerythrin subunits, were separated and analyzed using MALDI MS/MS. The attachment site was elucidated after interpretation of tandem MS spectra obtained from digests of PEB-peptide complexes. Figure 4.1.A shows m/z 800-1050 of the MALDI mass spectrum of a tryptic digest of the C-phycoerythrin alpha subunit complex. This work demonstrates the in-vitro attachment of PEB to the C-phycoerythrin alpha subunit complex using a newly developed enzyme complex CpeY and CpeZ. The peak at m/z 935 was identified as a PEB-tripeptide complex. The peptide sequence was deduced to be (R) C^{82}AR (D) wherein the PEB chromophore is covalently attached to the cysteine. Figure 4.2.A shows the region m/z 1100-1300 of the MALDI MS spectrum of a tryptic digest of C-phycoerythrin beta subunit after *in vitro* attachment of PEB using a newly developed enzyme, CpeS [16]. The peak detected at m/z 1250 was identified as a tryptic peptide complex containing a covalently bound PEB ligand. The sequence of this peptide was deduced to be (R) MAAC^{80} LR (D). Figure 4.3.A shows the m/z 960-1120 region of the mass spectrum of a tryptic digest of C-phycoerythrin alpha subunit and the peak at m/z 1089 was identified as a tryptic peptide covalently linked to PEB.

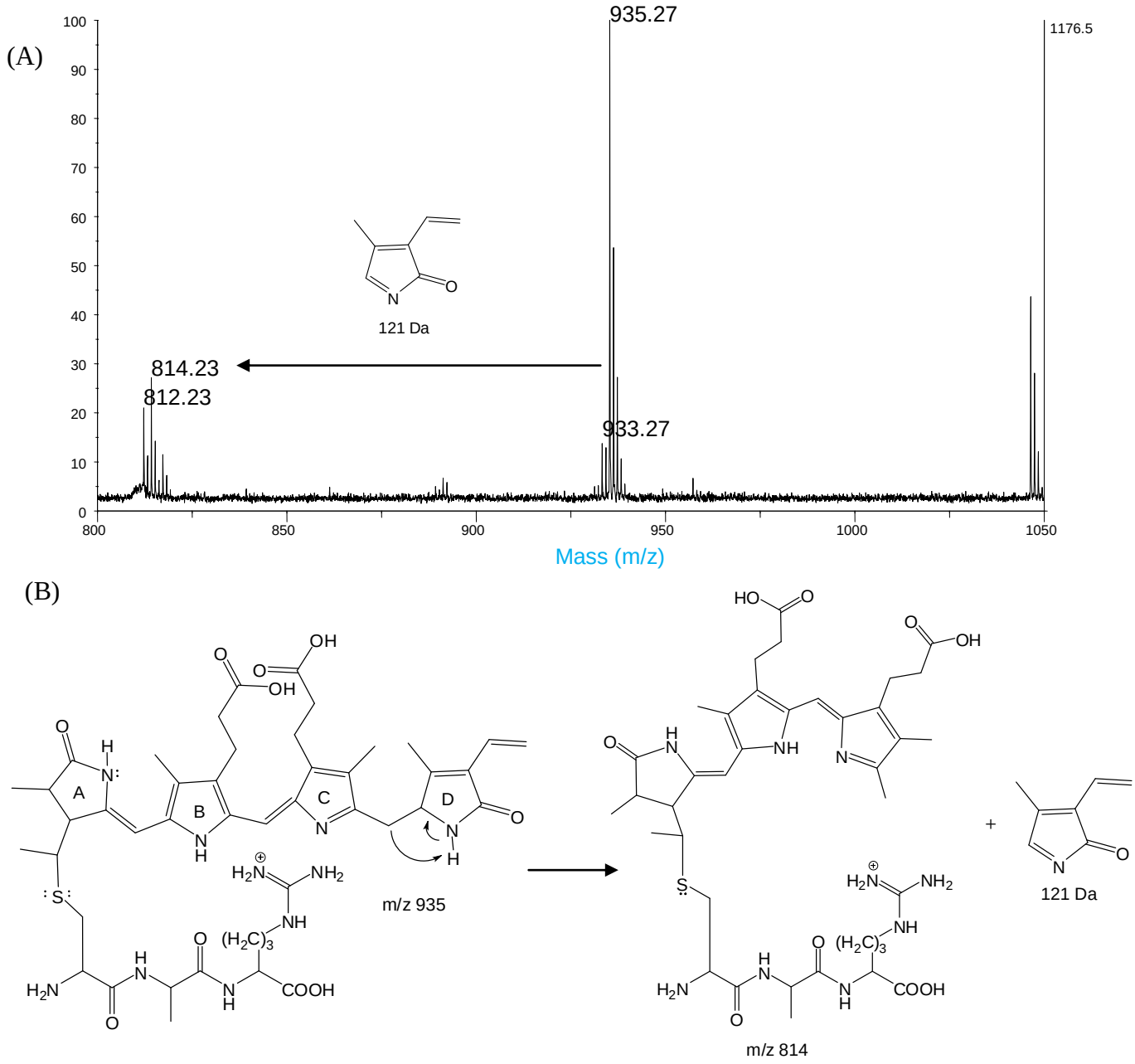


Figure 4.1. (A) Zoom into mass spectrum of CpeA tryptic digest MALDI MS spectrum. The peak at m/z 814 was identified as resulting from the heterolytic cleavage of the covalent bond between C15 – C16. (B) Proposed mechanism for the D-ring loss: the resulting ion detected at m/z 814 as well as the departing neutral are stabilized by Π-conjugation.

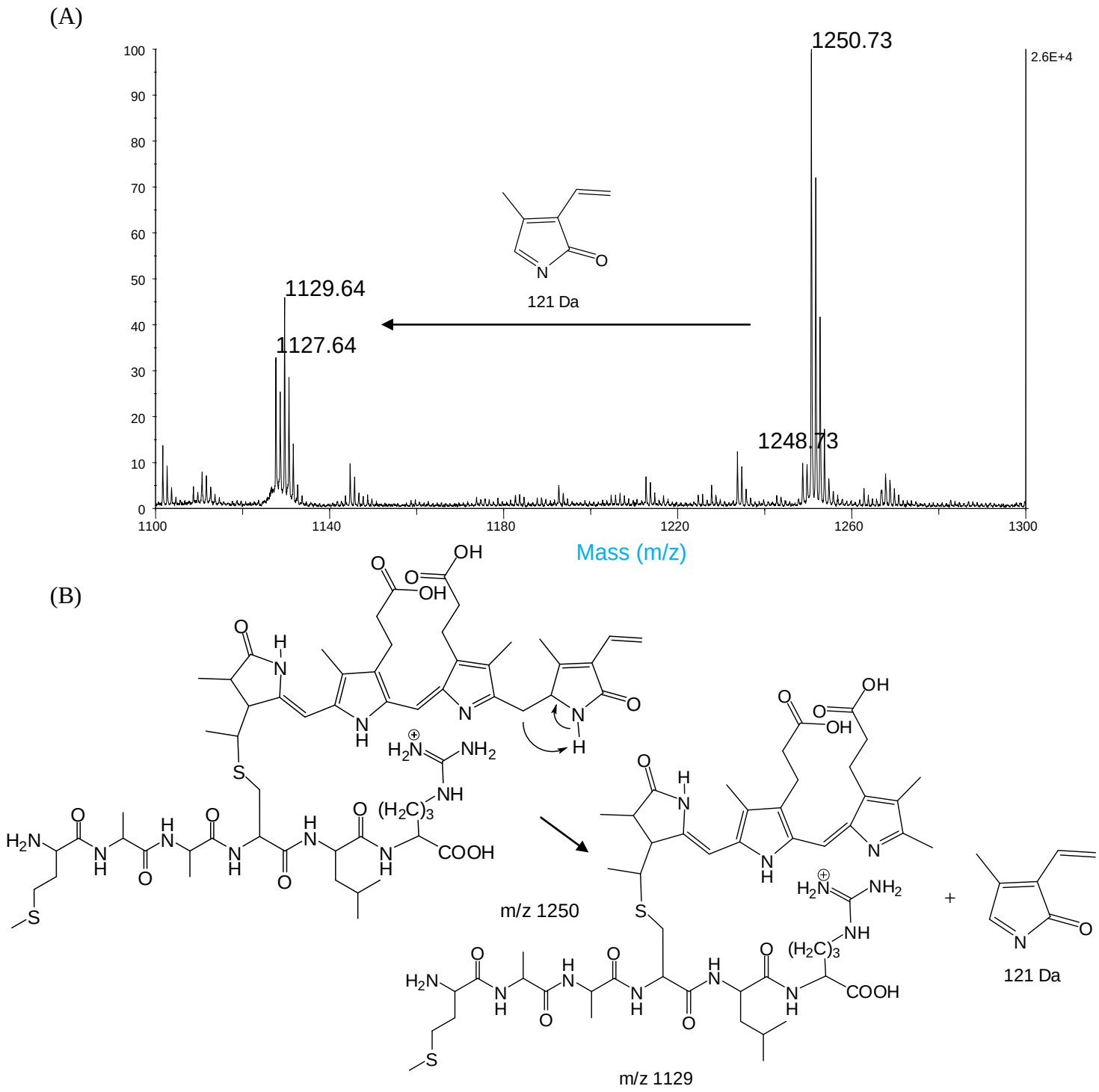


Figure 4.2. (A) MALDI mass spectrum of tryptic digest of CpeB. The peak at m/z 1129 was identified as resulting from the heterolytic cleavage of the covalent bond between C15 – C16. (B) Proposed mechanism for the D-ring loss: the resulting ion detected at m/z 1129 as well as the departing neutral are stabilized by Π-conjugation**.**

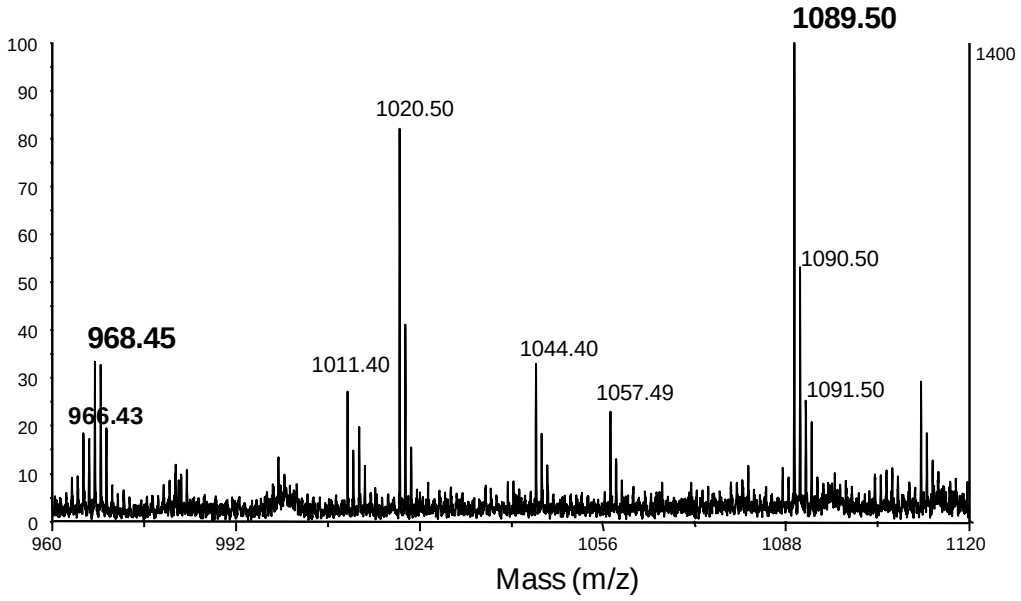


Figure 4.3. Chromopeptide from tryptic digest of CpeA-PEB complex. PEB attachment was catalyzed by a novel enzyme CpeY and CpeZ. The peak at m/z 1089 has been identified as a chromopeptide containing a PEB attached to Cys^{139}. Neutral loss of 121 Da leading to the detection of the peak at m/z 968 is detected again

When reviewing each of the mass spectra appearing in Figs. 4.1-4.3, it was noticed that, in addition to the peak corresponding to the PEB-tryptic peptide complex, in each case there is consistently an accompanying peak at 121 Da below the mass of the complex. That is, the peaks at m/z 814 in figure 1, m/z 1129 in figure 2 and m/z 968 in figure 3 are each detected at 121 Da below the mass of their respective PEB-complexes: m/z 935, m/z 1250 and m/z 1089. After careful inspection of tandem MS spectra of all the PEB-tryptic peptide complexes.[16] The additional peaks were deduced to result from the loss of the D-ring of PEB, which corresponds to a neutral loss of 121 Da. Figures 1(B), 2(B) show the reaction leading to the loss of the D-ring, which is a heterolytic cleavage of the C^{15}-C^{16} covalent bond of PEB. Notably, this cleavage occurs at a carbon-carbon bond that lacks double bond character, unlike the majority of resonance stabilized bonds comprising the PEB backbone. In addition to the relative weakness of this bond, the formed neutral is stabilized by π-conjugation.

In-order to confirm the structures of the peaks resulting from the neutral loss, tandem MS experiments were performed on each of the precursors which had been observed to lose 121 Da neutrals in the ion source. Figure 4.4 compares tandem MS spectra of m/z 935 and m/z 814 precursors from the same sample. Both tandem MS spectra contain a peak at m/z 349 which corresponds to the protonated peptide whose sequence was deduced to be (R) C^{82} AR (D). However, the peak at m/z 587 (assigned as protonated PEB) in the product ion spectrum of precursor m/z 935 (figure 4(A)), has been shifted to m/z 464, which is now assigned as the protonated tripyrrole-ring, which was formed from PEB upon loss of its D-ring. Analogous results were observed in comparing product ion mass spectra of precursor ions at m/z 1250 (Fig. 4.2) and 1089 (Fig. 4.3) with the corresponding peaks detected at m/z 1129 and 968, respectively (results not shown). These results provide evidence that the discussed peaks, each detected at 121 Da below the respective precursor, each result from the loss of the PEB D-ring.

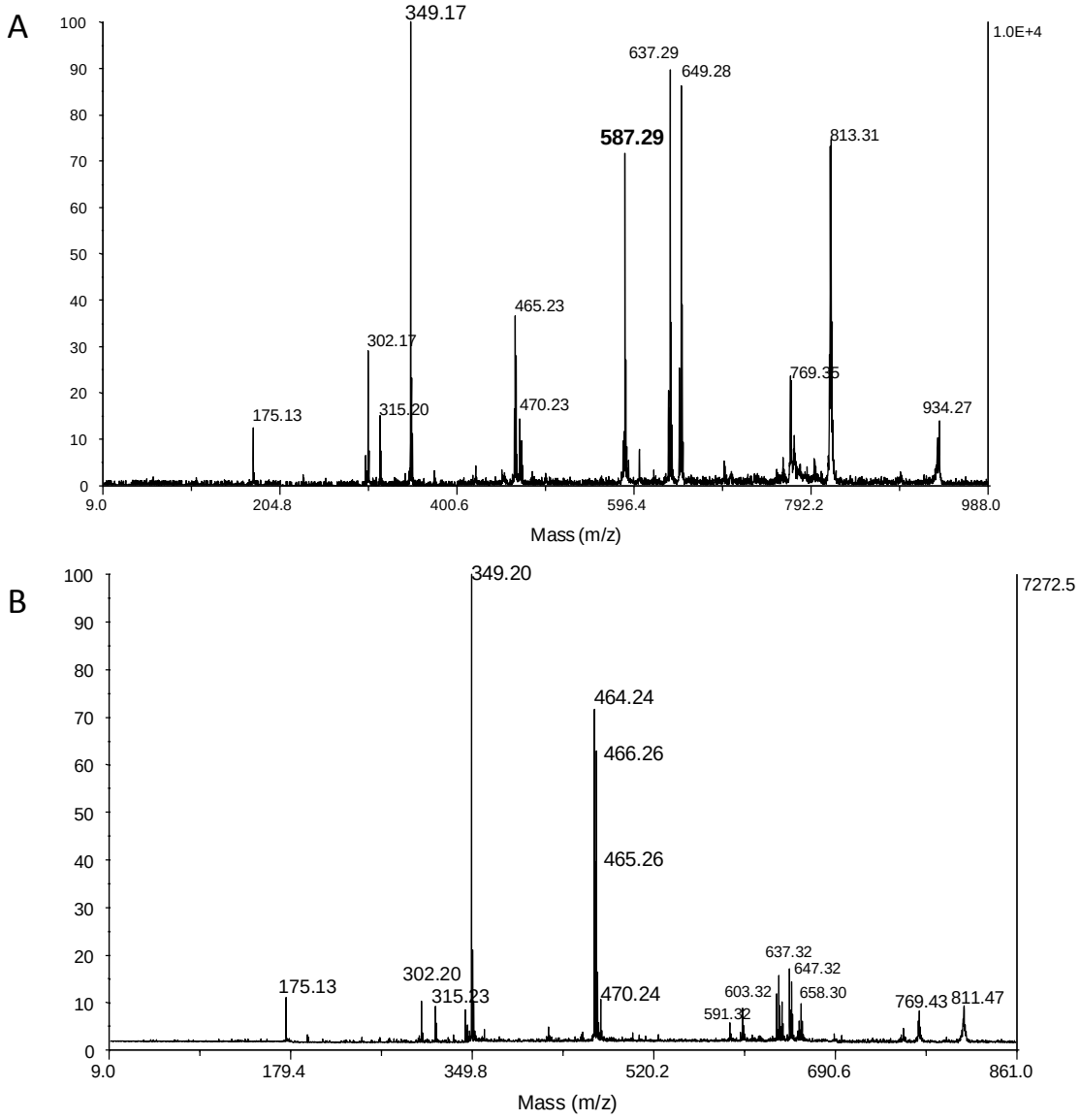


Figure 4.4. Tandem mass spectrum of precursor ion (A) m/z 935 and (B) m/z 814. Notice that the fragment at m/z 587 corresponding to protonated PEB (tetrapyrrole) has been replaced by a fragment at m/z 464, 465 and 466, which are various mesomers of a tripyrrole ring. Tripyrrol attached chromopeptide are therefore present in the MALDI MS spectrum.

In a follow-up experiment, holo-PE was isolated from *F. diplosiphon* and was subsequently subjected to tryptic digestion, with the goal of obtaining MALDI mass spectra of *in vivo* attached PEB complexes. *In vivo* attachment has been previously proven to proceed through A-ring attachment of PEB chromophores, thus, analysis of these samples with known structures enables us to more readily comprehend obtained mass spectra.

Figure 4.5 shows the MALDI mass spectra of three tryptic peptides containing PEB attached to: a) α-Cys^{82} b) β-Cys^{80} and c) α-Cys^{139} from the holo-phycoerythrin digest. Relative to each intact protonated peptide holding the PEB ligand, strong signals corresponding to the neutral loss of 121 Da were again detected in each case. Because the PEB configuration is known, i.e., *in vitro* attachment using the new bilin lyase exclusively links the PEB through its A-ring to the peptide, then the peak resulting from neutral loss of 121 Da can be unequivocally assigned as a D-ring loss. Moreover, we introduce the concept that the appearance of this 121 Da loss in the conventional MALDI mass spectrum is diagnostic of the fact that indeed PEB is singly linked to the peptide through the A-ring. If PEB were doubly linked through the A and D rings, or singly linked through the D ring, loss of the 121 Da neutral would be heavily disfavored.

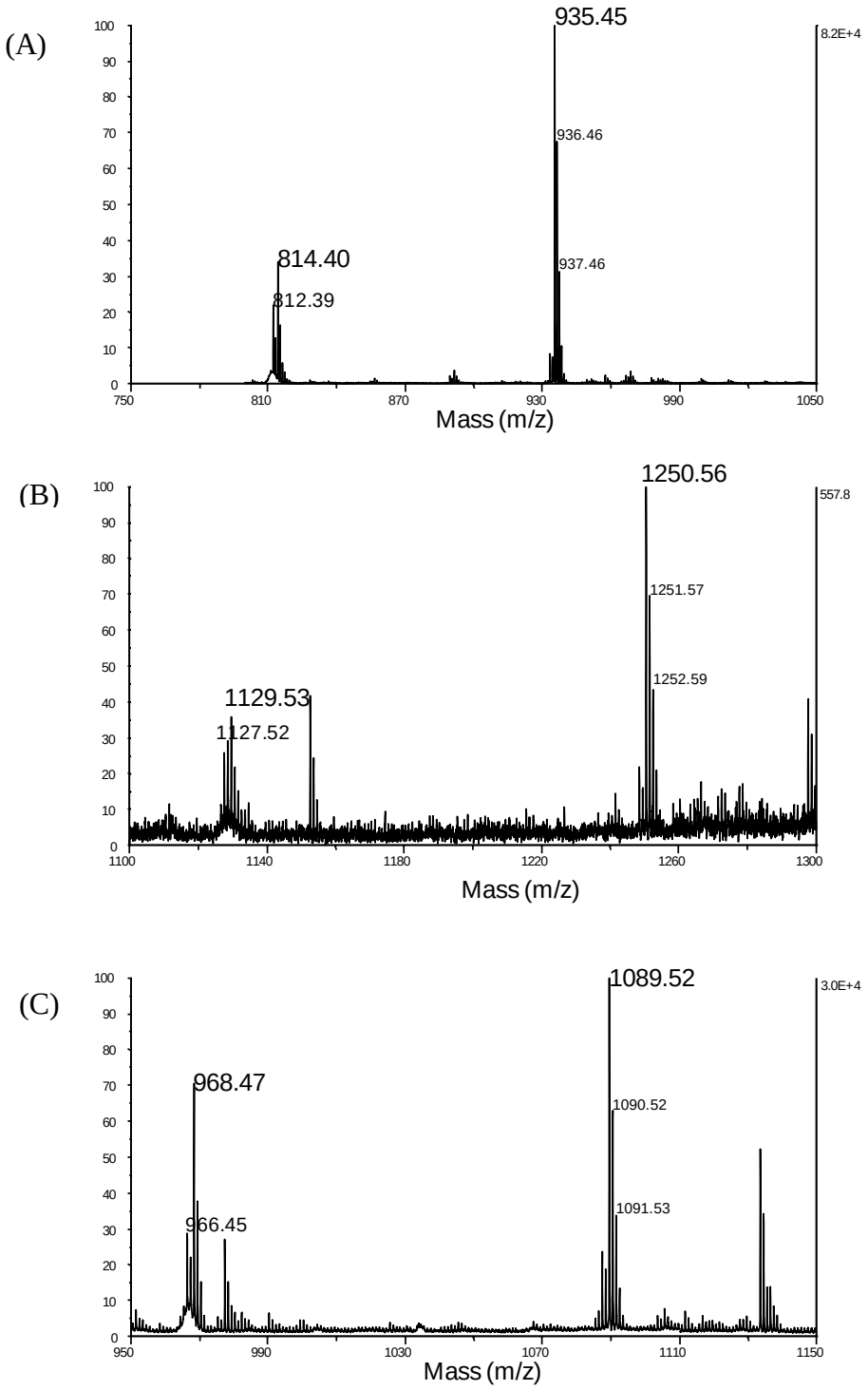


Figure 4.5. MALDI mass spectrum of the tryptic digest of Holo-Phycoerythrin. (A) chromopeptide complex PEB-tripeptide (R) CAR (D) PEB is attached through Cys^{82}. (B) Chromopeptide resulting from tryptic digest of CpeB. The peptide is (R) MAACLR (D) and it is attached to PEB through Cys^{80}. (C) Chromopeptide resulting from tryptic digest of CpeA, the peptide sequence is (R) GCAPR (D) and PEB is attached to Cys^{139}. The peak resulting from the neutral loss of 121 Da is detected in each experiment.

MALDI is known to be a soft ionization method wherein the matrix absorbs the brunt of the laser energy thereby protecting fragile analyte molecules from thermal decomposition. This also implies that little internal energy is taken up by (intact) desorbed species. Hence, few fragments are produced. The absorbance spectrum of a PEB-chromophorylated peptide was published [8] and showed that, aside from a main absorption band in the visible at 550 nm, the chromopeptide also absorbed between 350-370 nm in the near UV range. Therefore, we propose that laser photons emitted at 355 nm are directly absorbed by the PEB chromophore during MALDI irradiation, and that it is from this direct photon absorption that heterolytic cleavage of the C^{15}-C^{16} bond is induced. This means that efficient photo-induced direct cleavage of this bond can precede desorption of the tripyrrole- peptide complex into the gas phase.

We have thus established a diagnostic MALDI MS method, based upon the reproducible gas-phase behavior of PEB-linked tryptic peptides, for determining if bilin attachment (either single or double) to the D-ring has occurred. Upon laser irradiation, the PEB chromophore appears to directly absorb laser photons which, in the absence of attachment to the D-ring, results in the cleavage of the C^{15}-C^{16} bond and the formation of a tripyrrole (now fully conjugated)-plus-peptide complex with loss of a stabilized neutral.

When investigating the site-specificity for PEB attachment of a novel enzyme, which was thought to promote double attachment of PEB (A- and D- rings) to Cys^{48}-Cys^{59}, a complex was detected at m/z 4626.9 which was assigned as a PEB-tryptic peptide [Leu^{36}-Arg^{75}] containing both Cys^{48} and Cys^{59}. Figure 4.6 shows the MALDI mass spectrum of the complex. A peak at m/z 4040 was detected as well and was assigned as the free tryptic peptide [Leu^{36}-Arg^{75}].

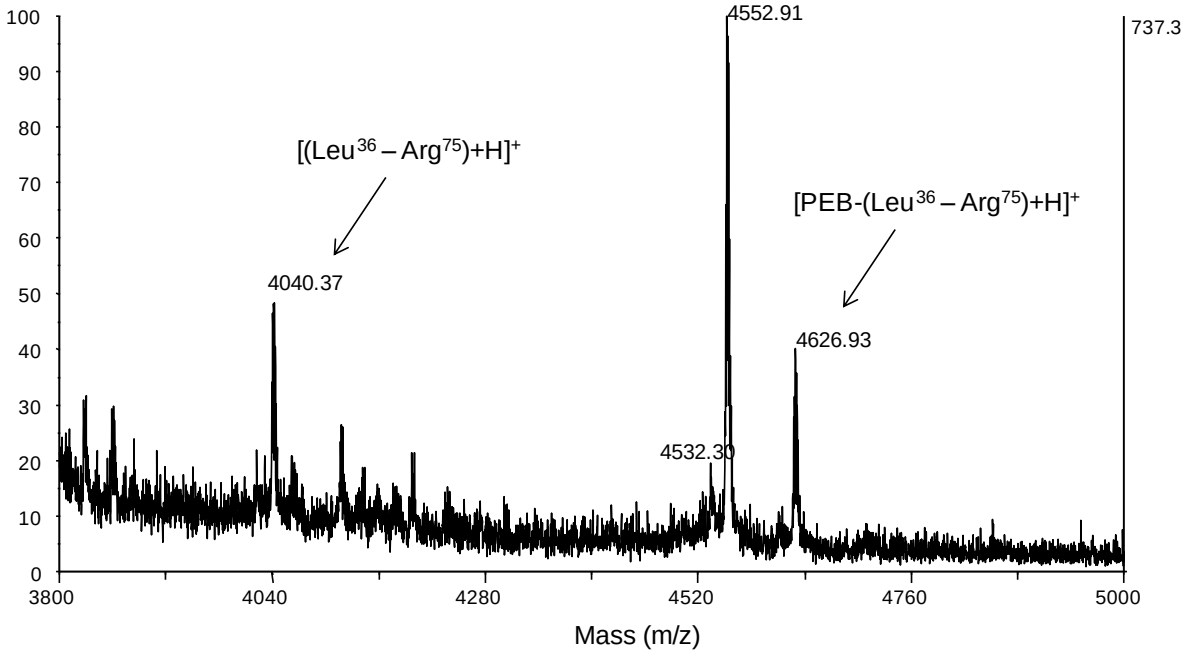
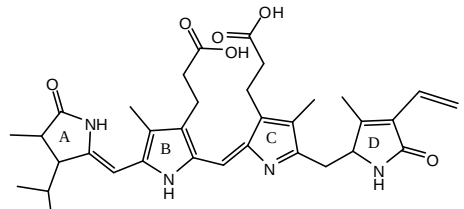


Figure 4.6. Mass spectrum of a chromopeptide detected at m/z 4627 resulting from tryptic digest of C-Phycoerythrin beta subunit. The peak at m/z 4040 was attributed to the free peptide, $[(Leu^{36}\text{-}Arg^{75})+H]^+$, and containing two cyteines: Cys 48 and Cys 59. No peak resulting from the neutral loss of 121 Da was detected.

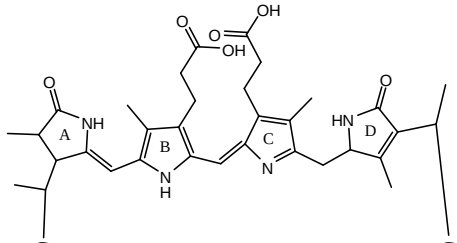
(A)



m/z 4627

Leu-Asp-Ala-Val-Asn-Ala-Ile-Ala-Ser-Asn-Ala-Ser-**Cys**-Met-Val-Ser-Asp-Ala-Val-Ala-Gly-Met-Ile-**Cys**-Glu-Asn-Gln-Gly-Leu-Ile-Gln-Ala-Gly-Gly-Asn-Cys-Tyr-Pro-Asn-Arg⊕

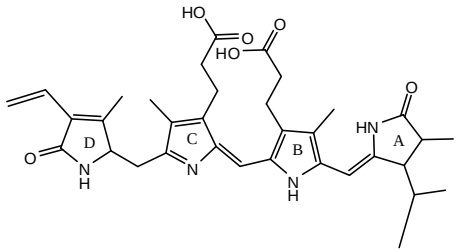
(B)



m/z 4627

Leu-Asp-Ala-Val-Asn-Ala-Ile-Ala-Ser-Asn-Ala-Ser-**Cys**-Met-Val-Ser-Asp-Ala-Val-Ala-Gly-Met-Ile-**Cys**-Glu-Asn-Gln-Gly-Leu-Ile-Gln-Ala-Gly-Gly-Asn-Cys-Tyr-Pro-Asn-Arg⊕

(C)



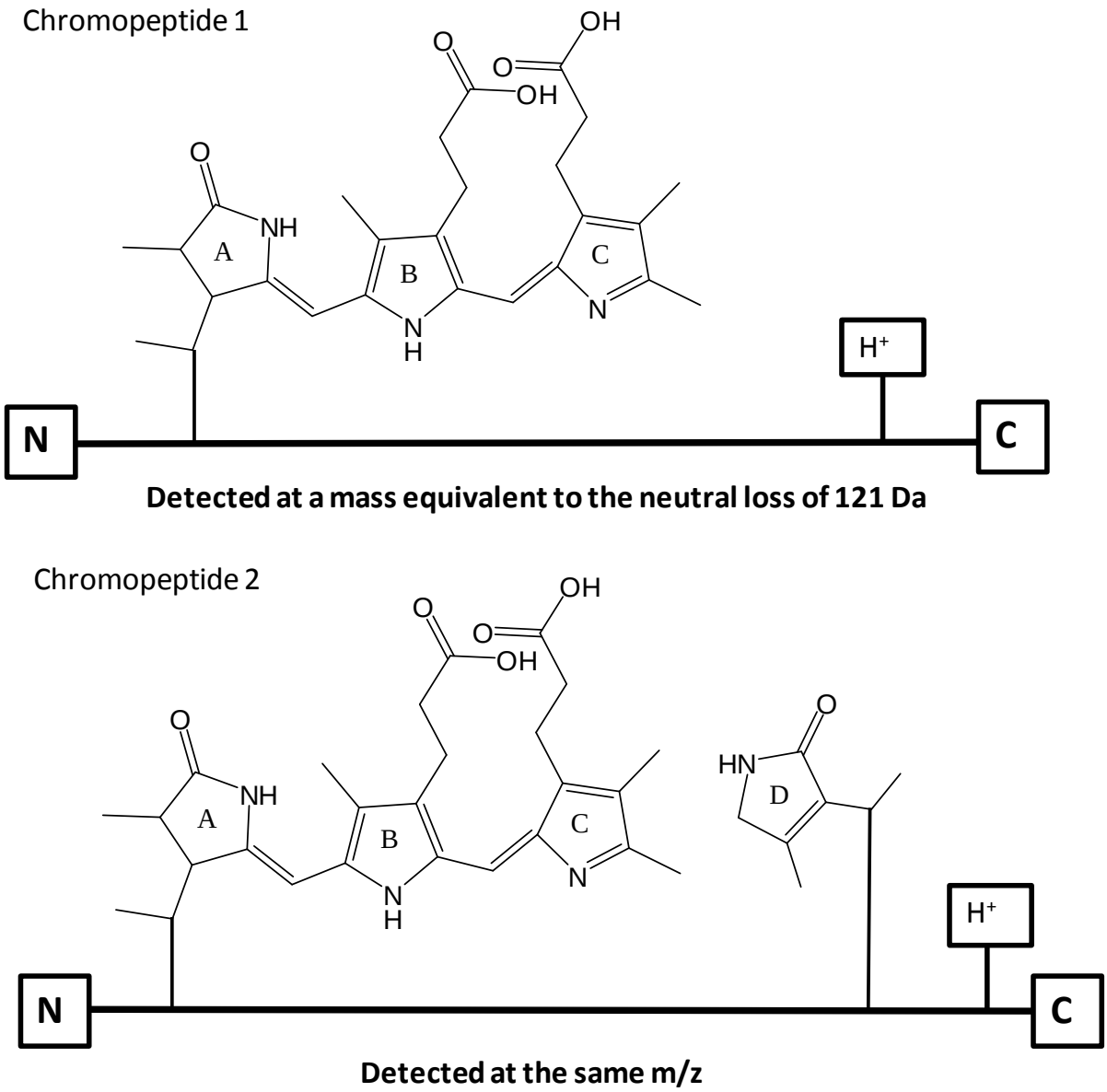
m/z 4627

Leu-Asp-Ala-Val-Asn-Ala-Ile-Ala-Ser-Asn-Ala-Ser-**Cys**-Met-Val-Ser-Asp-Ala-Val-Ala-Gly-Met-Ile-**Cys**-Glu-Asn-Gln-Gly-Leu-Ile-Gln-Ala-Gly-Gly-Asn-Cys-Tyr-Pro-Asn-Arg⊕

Scheme 4.3. Chromopeptide resulting from tryptic digest of C-phycoerythrin beta subunit. The peptide starts from Leu^{36} until Arg^{75}. The peptide, in holo-PE, contains a doubly attached PEB chromophore at both Cys^{48} and Cys^{59}. When performing PEB attachment experiments, probing the attachment at position Cys^{48} and Cys^{59} is impossible because the singly and doubly attached PEB complexes have the same mass. Tandem MS experiments at such high masses are very inefficient and result in few fragments.

However, *a priori*, the question as to whether the PEB chromophore was doubly attached (implicating both cysteines) or singly attached to either cysteine is not readily answerable because the resulting PEB-tryptic peptide complexes formed in the above three cases would be detected at the same mass. Scheme 3 shows the structure of the resulting complexes and Schemes 4.2 and 4.4 show schematic diagrams of a more general issue that investigators could be facing: when a PEB chromophore is attached to cysteines belonging to the same tryptic peptide, the complex formed in the case of a singly or a doubly linked PEB will have the same mass, making it impossible to fully characterize the attachment using MALDI MS only. Moreover, due to the rather high mass of the resulting chromopeptide, collision induced dissociation experiments are not efficient and they produce few fragments, if any.

According to our results appearing in Figures 4.1-4.3 and 4.5, singly-attached PEB chromophores, when analyzed using MALDI, will undergo a gas-phase fragmentation leading to the loss of the D-ring. In all instances where single attachment at the A-ring is expected, the loss of a 121 Da neutral was observed. However, no such loss (that would have produced m/z 4505.9) was observed in the MALDI mass spectrum shown in figure 7. We rationalize the fact that this peak is missing by considering that the bilin D-ring is covalently attached to a cysteine through a thioether bond at position 48 or 59 (Scheme 4.3.B) Thus, even if the C^{15}-C^{16} bond undergoes a heterolytic cleavage as the result of laser irradiation, no 121 Da loss will be observed because the D-ring is covalently held in place (Scheme 4.4). On the other hand, if PEB was singly attached through its A-ring to the peptide, then a loss of 121 Da would have been expected. Therefore, the absence of a peak resulting from the neutral loss of the D-ring provides evidence that the PEB is indeed “doubly” attached.



Scheme 4.4. According to our previous results, the C(15)-C(16) bond undergoes a heterolytic cleavage in each chromopeptide. In which case, chromopeptide 1 will have a peak detected corresponding to the neutral loss of the D-ring. But, since the D-ring is covalently attached in chromopeptide 2, the bond cleavage will not result in detection of a peak corresponding to the neutral loss.

An alternative means to establish that a double linkage exists between PEB and a peptide would be to use an enzyme other than trypsin for the protein digestion. If digestion produced a cleavage of the peptide backbone in between each cysteine, then the PEB that was initially doubly attached should provide a fragment that links the two cysteines. The use of 1D and 2D NMR could be an option as well, but this technique is far less sensitive than MALDI MS and the spectral interpretation will be more time consuming. A single MALDI MS spectrum and a simple interpretation method can confirm the type of attachment that is involved.

4.5 Conclusions

We report in this work an easy, straightforward and accurate methodology in order to discriminate between doubly- and singly-attached PEB. When PEB is attached to a tryptic peptide through its A-ring, the C^{15}-C^{16} bond undergoes a heterolytic cleavage upon absorption of laser photons (λ=355nm) which results in the loss of the D-ring. This cleavage produces a peak that is 121 Da below the mass of the protonated chromopeptide complex. The photophysical property described here could be used in order to identify the number of linkages that exist between a PEB chromophore and a tryptic peptide containing two potential sites of attachment. Singly and doubly-attached PEB-peptides will have the same mass. However, a singly attached PEB will result in a peak detected at 121 Da below the mass of the complex. Failure to detect such peak is a strong evidence of double attachment.

4.6 References

[1] Gantt, E., Structure and function of phycobilisomes: light harvesting pigment complexes in red and blue green algae. *Inernational Review of Cytology.* **1980**, 66, 45-80

[2] Glazer, A. N., Light Guides: directional energy transfer in a photosynthetic antenna. *J. Biol. Chem.* **1989**, *264*, 1-4

[3] Scheer, H.; Zhao, K. –H. Biliprotein maturation: the chromophore attachment. *Molecular Microbiology*. 2008, 68, 263-276

[4] Scheer, H., Biliproteins. *Angewandt Chemie International Edition in English.***1981**, 20, 241-261

[5] Shirmer, T.;Bode, W.; Huber, R., Refined three-dimensional structures of two cyanobacterial C-phycocyanins at 2.1 and 2.5A resolution: A common principle of phycobilin-protein interaction. *Journal of Molecular Biology* **1987**, 196, 677-695

[6] Duerring, M.; Schmidt, G. B.; Huber,R.; Isolation, crystallization, crystal structure analysis and refinement of constitutive C-phycocyanin from the chromatically adapting cyanobacterium *Fremyella diplosiphon* at 1.66A resolution. *Journal of Molecular Biology* **1991**, 217, 577-592

[7] Schoenleber, R. W.; Lundell, D. J.; Glazer, A. N.; Rapoport, H., Bilin Attachment sites in the α and β Subunits of B-phycoerythrin. Structural studies on the singly linked Phycoerythrobilins. *The Journal of biological Chemistry*. **1984**, 259,5485-5489

[8] Schoenleber, R. W.; Leung, S. L.; Lundell, D. J.; Glazer, A. N.; Rapoport, H., Chromopeptides from Phycoerythrins. Structure and Linkage of Phycoerythrobilin Tryptic Tripeptide Derived from a B-Phycoerythrin. *The Journal of the American Chemical Society.***1983**, 105, 4072-4076

[9] Lagarias, J. C.; Klotz, A. V.; Dallas, J.; Glazer, A. N.; Bishop, J. E.;O'connell, J. F.; Rapoport, H., Exclusive A-ring Linkage forSingly Attached Phycocyanobilins and Phycoerythrobilins in Phycobiliproteins. Absence of singly D-ring-linked Bilin. *The Journal of Biological Chemistry.***1988,** 263, 12977-12985

[10] Bishop, J. E.; Lagarias, J. C.; Nagys, J. O.; Schoenelber, R. W.; Rapoport, H.; Klotz, A. V.; Glazer, A. N., Phycobiliprotein-Bilin Linkage Diversity.Structural study on A- and D-ring-linked Phycocyanobilins. *The Journal of Biological Chemistry*. **1986**, 261, 6790-6796

[11] Klotz, A. V.; Glazer, A. N.; Bishop, J. E.; Nagys, J. O.; Rapoport, H.; Phycobiliprotein-Bilin Linkage Diversity.Structural study on A- and D-ring-linked Phycerythrobilin. *The Journal of Biological Chemistry*. **1986**, 261, 6797-6805

[12] Schoenleber, R. W.; Lundell, D. J.; Glazer, A. N.; Rapoport, H., Bilin Attachment sites in the α and β Subunits of B-phycoerythrin. Structural studies on the singly linked Phycoerythrobilins. *The Journal of biological Chemistry*. **1984**, 259,5481-5484

[13] Wilbanks, S. M., Wedemayer, G. J., Glazer, A. N. Post-translational Modifications of the β Subunit of a Cryptomonad Phycoerythrin. sites of Bilin Attachment and Asparagine Methylation. *The Journal of Biological Chemistry*. **1989**, 264, 17860-17867

[14] Wilk, K. E.; Harrop, S. J.; Jankova, L.; Edler, D.; Keenan, G.; Sharples, F.; Hiller, R. G.; Curmi, P. M. G. Evolution of a light-harvesting protein by addition of new subunits and

rearrangement of conserved elements: Crystal structure of a cryptophytes phycoerythrin at 1.63-Å resolution. *Proceedings of the National Academy of Science* **1999**, 96, 8901-8906

[15] Wiethaus, J., Busch, A. W. U., Kock, K., Leichert, L. I., Herrmann, C., and Frankenberg-Dinkel, N. CpeS Is a Lyase Specific for Attachment of 3Z-PEB to Cys-82of β-phycoerythrin from *Prochlorococcus marinus* MED4, *The Journal of Biological Chemistry.* **2010,** 285, 37561-37569

[16] Biswas, A.; Boutaghou, M. N.; Alvey, R. M.; Kronfel, C. M.; Cole, R. B.; Bryant, D. A.; Schluchter, W. M. Characterization of the Activities of the CpeY, CpeZ, and CpeS Bilin Lyases in Phycoerythrin Biosynthesis in Fremyella diplosiphon Strain UTEX 481. *Journal of Biological Chemistry*. **2011**, 286, 35509-35521

[17] Isailovic, D., Sultana, I., Phillips, G. J., and Yeung, E. S. Isolation and characterization of R-phycoerythrin subunits and enzymatic digests. *Journal of Chromatography A.* **2004**, 1051,119-130

[18] Bennett, A.; Bogorad, L. Properties of Subunits and Aggregates of Blue-Green Algal Biliproteins. *Biochemistry* **1971**, *10*, 3625-3634

[19] Arciero, D. M.; Bryant, D. A.; Glazer, A. N.; *In Vitro* Attachment of Bilins to Apophycocyanins. Specific covalent adduct formation at cysteinyl residues involved in phycocyanobilin binding in C-phycocyanin. *The Journal of Biological Chemistry*. **1988**, 263, 18343-18349

Chapter 5

Evaluating the false discovery rate for a *Fundulus grandis* proteome project

5.1 Abstract

This chapter assesses the accuracy of protein identifications reported in a large-scale proteome expression project. Specifically, we sought to quantify the rate of false protein identifications when investigating the protein expression profiles in five tissues of the Gulf killifish, *Fundulus grandis*. MALDI MS and MS/MS spectra were submitted to a "target-decoy" search, in which spectra were compared against real and randomized protein sequences. Of 864 protein samples examined, 405 returned matches against the real protein database and two returned matches against the randomized protein database. Assuming that two random matches also occurred in the searches against the real protein database, a total of 4 out of 864 searches returned incorrect matches (0.46%). If expressed as a proportion of the matches against the real database (4/405), the false positive identification rate is still less than 1%. Both measures are lower than the P value normally accepted for statistical significance in scientific hypothesis testing (5%) and support the view that the vast majority (>99%) of the protein identifications based upon automated database searching of MS and MS/MS spectra are accurate.

5.2 Introduction

Proteomics is defined as the qualitative and quantitative analysis of the protein expression profile of a cell, tissue or organism under a defined set of conditions [1-4]. The field of proteomics has drawn great attention since full genomes, such as the human genome, were sequenced [5,6]. Once genome sequences became available, research efforts shifted toward the product of the gene expression: proteins.

At the same time as genomes were being sequenced, advances in mass spectrometry increased its application to biological molecules, including proteins. The advent of soft-ionization techniques such as Electrospray Ionization (ESI) [7] and Matrix Assisted Laser Desorption Ionization (MALDI) [8] propelled mass spectrometry as a technique of choice for protein analysis [9-11]. Through mass spectrometry, the analysis of intact proteins and peptides without inducing unwanted fragmentation was made possible. These features along with improvements in sensitivity and automation of mass spectrometer analyzers enabled high throughput proteome analysis.

A central challenge in proteomic analysis is the variety and complexity of the collection of protein expressed by a given tissue at any particular time. Tissue extracts generally include thousands of proteins that vary by size, solubility, and isoelectric point (pI). This complexity requires means of separating proteins prior to mass spectrometry analysis. Liquid chromatography (LC) and gel electrophoresis (GE) are the main separation techniques in proteomics experiments [12,13]. Liquid chromatography is generally performed on proteins that have been cleaved using a site-specific protease. Gel electrophoresis is usually applied to undigested protein mixtures, with resolved proteins digested after separation [14]. Peptides resulting from proteolysis are analyzed using mass spectrometry. Peptide mixtures are submitted to mass spectrometric analysis to determine mass to charge ratio (m/z) for each

peptide. Subsequent fragmentation of selected peptides yields information about their amino-acid sequence.

Protein identification proceeds on two levels [15,16]. The first level of identification is known as Peptide Mass Fingerprinting, PMF [14, 17]. The MS spectra of the unknown peptides are compared to theoretical MS spectra of in-silico generated proteolytic peptides. The second level of identification is known as Peptide Fragment Fingerprinting, PFF, when a peptide tandem mass spectrum is compared to the spectrum of the fragmented peptide it was matched with. The two levels of identification are complementary and the accuracy of an identification increases when positive matches are recorded at each level.

Due to the volume of data to be processed, identifications are performed by sequence database search engines. These search engines compare experimental mass spectra and theoretical ones. Experimentally-derived mass spectra yield lists containing masses and intensities. Search engines compare these data to in-silico digested proteins from a designated protein database. Protein matches are then assigned a score that represents the degree of confidence that the identification is not simply due to random chance. Several database searching tools are available: Sequest™ [18] ProteinProspector™ [19] and MASCOT™ [20].

The MASCOT database search engine was developed from an early scoring algorithm known as MOWSE [21]: Molecular Weight Search. MASCOT relies on a probability-based scoring strategy. The algorithm will calculate the probability that a match between an experimental set of data and a sequence in the database is due to random chance. This probability is calculated for all of the possible combinations of matches between the experimental data and the database sequences. Any event that has a lower than 5% probability to be random is considered a match. The match that has the lowest probability of being random will be given the highest score. The probability of a random match increases when the size of the database increases. As a result, MASCOT scoring algorithm increases

the threshold score of an identification when the size of the database increases. The threshold score (S) is related to probability (P) and the number of sequences in the database (k) by the formula:

$$S = -10 \times \mathrm{Log}_{10}(P/k)$$

With any large-scale, database searching approach that might include thousands or even millions of individual searches, false positive identification is a major concern [22]. A false positive identification occurs when a match of MS data returns a score equal to or greater than the threshold, ranks first amongst all possible matches, but is due to random chance and is an incorrect assignment. An application within MASCOT was developed to assess the scoring accuracy for a given database searching results. The "target-decoy search" script creates a "randomized sequence database", in which the amino-acid sequences of all proteins in the database are randomized creating new protein sequences that have the same molecular weights. If a PMF or PFF matches a randomized sequence better than a true sequence, then this is considered a false positive. The number of random matches in a large-scale experiment is then used to determine the false discovery rate (FDR), which in turn is a gauge of reliability of the protein identifications. Moreover, editorial boards of a number of journals are requiring investigators to provide a measure of identification certainty by calculating an estimate of the false-positive rate or false discovery rate (FDR) when protein and peptides identification are documented. [23]

Mass spectrometric-based proteomic approaches are being increasingly applied to assess the effects of environmental stress on the proteome in a variety of organisms [24, 25]. Comparison of the heat stress response between a warm-adapted mussel *Mytilus galloprovincialis* and a cold-adapted *M. trossulus*, showed species-specific changes in the abundance of proteins involved in molecular chaperoning, protein turnover, and energy metabolism [26]. The proteomic response to osmotic stress has been investigated in the

dogfish shark, *Squallus acanthias* [27]. Results showed that this elasmobranch reacts to changes in water salinity by increasing the synthesis of proteins involved in lysis, phosphorylation, and hydroxylation functions. When a rainbow trout hypodermal fibroblast cell line is exposed to long term anoxia, it was found that proteins involved in energy metabolism, such as isocitrate dehydrogenase, were down-regulated and proteins involved in cytoskeleton functions were up-regulated [28].

The application of MS-based proteomic approach in environmental biology, however, can be limited by the lack of a complete genome sequence for the organism under study. Among fish, z*ebrafish* (*Danio rerio*) [29-31] and puffer fish (*Takfugu rubripes*) genomes have been sequenced and protein databases are available (although not completely annotated). For species that do not have a genomic sequence, protein identification relies upon searches against protein databases of closely related species. These databases may not contain all the proteins belonging to the targeted organism. Moreover, even when homologous proteins are available, if they differ in amino acid sequence, they will generate peptides of differing mass that may prevent a match against experimentally determined masses. As a result, identification rates for organisms lacking genome sequences are usually low. Whether this is associated with an increased false discover rate is an open question.

Fundulus grandis occurs in coastal marshes of the northern Gulf of Mexico and has a broad tolerance for variation in salinity, temperature, and oxygen [32]. These features, combined with its availability and ease of laboratory maintenance makes *Fundulus grandis* a useful model for proteomic studies of fish. This chapter describes a protocol to assess the false positive rate for identifications obtained in a proteomic survey of *Fundulus grandis* tissue [33]. Because there is no genome database for this species, protein identifications were made by searching proteins from all of *Actinopterygii* (ray-finned fishes). Using the script provided by MASCOT, we created a randomized database and submitted MALDI MS and

MS/MS spectra to a search against this target-decoy database. In this chapter, we report the results of the *Fundulus grandis* proteome expression project as well as an evaluation of the accuracy of our results by providing a false discovery rate.

5.3 Materials and Methods

Acetonitrile, α-cyano-4-hydroxycinnamic acid, trifluoroacetic acid (TFA) were purchased from Sigma-Aldrich (St Louis, USA). *Fundulus grandis* were obtained from a local bait shop and kept in an aquarium until used.

Protein extraction, 2D gel electrophoresis, gel imaging, cutting and in-gel digestion were all protocols performed by Dr. Naga Abbaraju. More details about each protocol can be found in reference [33].

Protein extraction, 2D gel electrophoresis, gel imaging, cutting and in-gel digestion. Samples of 600 µg of *F. grandis* protein extract from five different tissues (liver, brain, heart, muscle and gill) were prepared for 2D gel electrophoresis. First dimension (IEF) was performed using commercially available immobilized pH gradient (IPG) strips (GE health care, Piscataway, NJ, USA) with Ettan IPGphor II Isoelectric Focusing Unit (GE health care, Piscataway, NJ, USA). Second dimension was SDS PAGE using 12.5% polyacrylamide gels according to Laemmli [34]. Gels were stained using modified Neuhoff's colloidal Coomassi protocol and de-stained in water. Gels were imaged with a GS 700 densitometer (BIORAD, Hercules, USA) and analyzed using PDQUEST ^{TM} 2D analysis software (BIORAD, Hercules, USA) for spot cutting. The most abundant spots from PDQUEST^{TM} 2D analysis software were excised from the gels using EXQuest^{TM} spotcutter (BIORAD, Hercules, USA). The gel spots were collected in 96-wellplates and trypsin digested using sequencing-grade trypsin (Promega, Madison, WI, USA) following the protocol from Investigator Progest^{TM} automatic digester (Genomic Solutions, MI, USA).

Mass Spectrometry analysis. Peptides resulting from in-gel digestion by trypsin were suspended in 20 µl of 50% acetonitrile. Two microliters of the tryptic peptide solution were mixed with an equal volume of α-cyano-4-hydroxycinnamic acid (10 mg ml^{-1} in 1:1 : acetonitrile : 0.1% aqueous TFA). Approximately 0.5 µL of each sample were spotted twice on a MALDI target for a total of 192 spots per 96-well plate. MALDI MS and tandem MS experiments were performed on an ABI 4800 MALDI TOF/TOF instrument (Foster City, CA), equipped with a Nd: YAG Laser (355nm, 20kHz). Peptide mass fingerprints were acquired using 400 laser shots in a mass range between m/z 800 and 4000. Two trypsin autolysis peaks at m/z 842 and 2211 were used as internal standards for MS calibration. The ten most intense peptide precursors were selected for MS/MS product ion acquisitions and the MS/MS spectra were acquired using 1000 laser shots.

Protein Database Searching. The MS spectra were transferred to GPS explorer software™ version 3.6 (Build 3.2) that used an underlying search algorithm of a locally installed copy of the MASCOT software programs, version 2.1 (Matrix Science; http://www.matrixscience.com). The data were searched against the Actinopterygii fishes subset of the non-redundant sequences deposited with NCBI database (updated as of March, 15th, 2011). Parameters for MASCOT database search were set as follows: fixed modification = carbamidomethylation of cysteine residues; variable modification = oxidation of methione residues; maximum missed cleavages = 1; PMF mass tolerance = 100 ppm; and MS/MS fragment tolerance = 0.5 Da. The spots identified as unknown proteins were submitted to a BLAST (Basic Local Alignment Search Tool) [35, 36] search to determine potential homologous proteins. De novo sequencing, provided by GPS explorer ™, was performed on spectra that did not return any match and the resulting sequences were submitted to a BLAST search.

Target-Decoy database searching. A target-decoy database was created according to MASCOT ^{TM} instructions [37]. A Perl script was downloaded from MASCOT and was executed using windows command prompt. The decoy database was created by randomizing the target database protein sequences and appended to the target database. Once the "target-decoy" database was created, it was downloaded to the MASCOT server. The "target-decoy" database searches were performed using the same specifications as our previous searches.

5.4 Result and discussion

MALDI mass spectrometry and tandem mass spectrometry experiments were performed to identify the most abundant proteins in 2D gels of *F. grandis* liver, brain, heart, gill, and muscle. Peptide mass fingerprinting (PMF) and peptide fragment fingerprinting (PFF) resulted in the identification of 405 proteins out of 864 gel spots submitted for analysis, which is equivalent to a 47% identification rate. A complete list of identified proteins per tissue is available in appendix tables 1-5. Individual proteins were identified by matching from one to fifty peptide masses, with an average of 9.7 peptides matched per identified proteins (fig. 5.1.A).

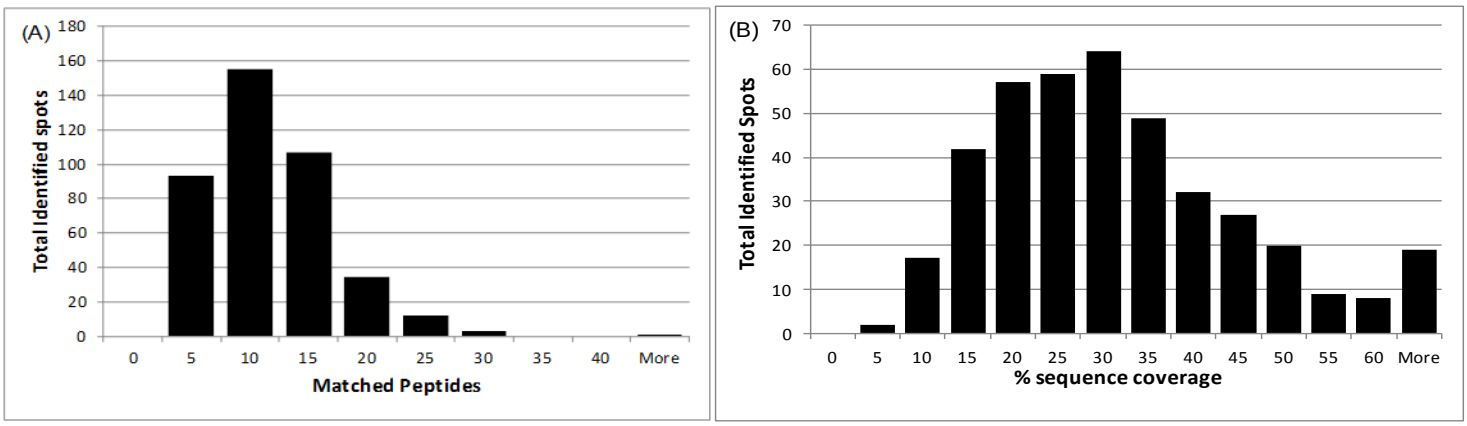


Figure 5.1. Frequency distribution histograms of peptides from identified proteins in all tissues of *Fundulus grandis*. Histogram (A) X- axis represents the number of matched peptides and Y-axis the number of identified spots. Histogram (B) X-axis represents the percent of sequence coverage and Y-axis the number of identified spots.

When introducing the MOWSE search tool, Papin *et al.* [38] reported that 99% of the proteins were identified correctly using five peptides or less, with a mean value of 3.6 peptides. The mean value of matched peptide per protein is higher in our experiments, which increases the confidence in our identifications. Moreover, the sequence coverage ranged from a minimum of 5% to a maximum of 83% with an average protein sequence coverage of

29.7% (fig.5.1.B). When proteins were identified only by PMF, the MASCOT score increased with the number of peptides matched. When specific peptide sequences yielded PFF that matched the database, the individual ion scores were included in the MASCOT score, meaning that fewer peptide matches were necessary to get a score above the threshold for a positive identification. The highest scores were obtained when a given protein had multiple peptides with PFF that matched a protein from the database. Scores attributed to the identified proteins varied from the minimum score of 67 to a maximum score of 747 (figure 5.2).

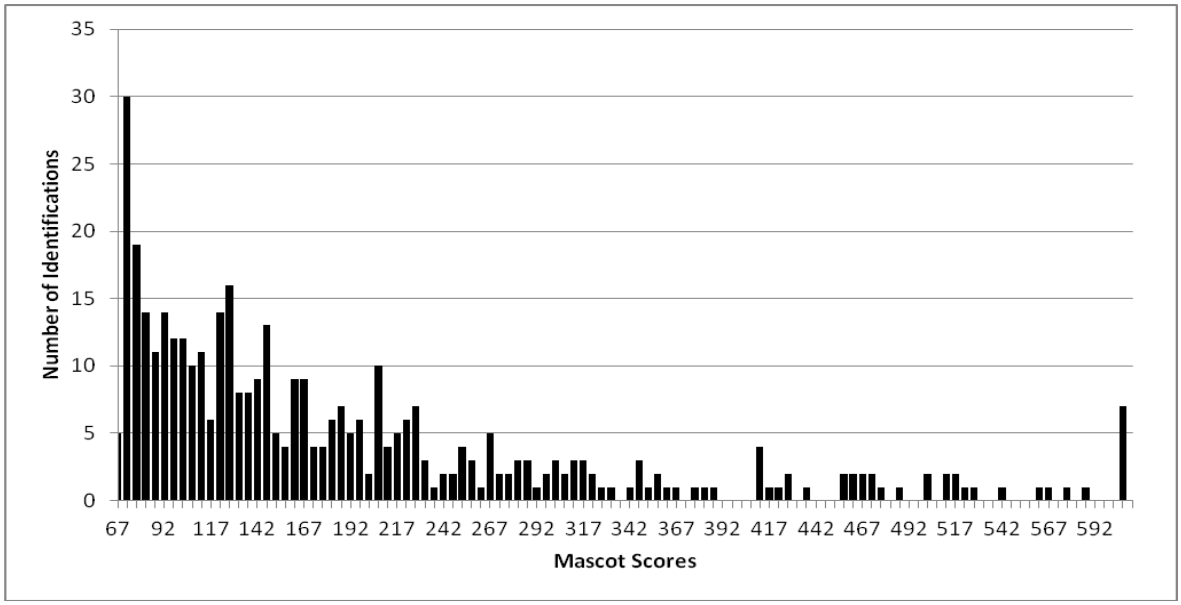


Figure 5.2. Histogram summarizing the scores of the 405 identified protein. Scores of identified proteins ranged from 67 to a maximum score of 757.

The ratio of identified protein decreased from plate 1 to plate 2 in all tissues as shown in table 5.1. Plate 1 contained the 96 most intense 2DE gel spots and plate 2 contained the next 96 most abundant proteins. Lower amounts of proteins in the scond plate leads to fewer and less abundant proteolytic peptides, the result was a lower chance of a protein match after database searching.

Tissue	Plate	Protein Assignments	% Identified	Decoy Hits
Liver	1	66	68.8	0
	2	27	28.1	1
Muscle	1	49	51.0	0
Brain	1	65	67.7	0
	2	35	36.5	0
Heart	1	64	66.7	0
	2	28	29.2	0
Gill	1	51	53.1	1
	2	20	20.8	0

Table 5.1. Result summary of the peptide mass and peptide fragment fingerprinting. Database searches were performed against the NCBI protein database of Actinopterygii (updated as of 03/2011). Plate 1 represents the 96 most abundant proteins and plate 2 represents the next 96 most abundant proteins.

As stated in the introduction, *Fundulus grandis* genome has not yet been reported. Therefore, the database searches were performed against an NCBI database of fishes that belong to the same class, namely Actinopterygii. Most of the proteins contained in this database belong to fishes whose genomes are sequenced, such as zebrafish and pufferfish. Approximately half of the trypsin-digested gel spots subjected to MS and database searching were not matched to this protein database. There are several reasons why these proteins were not matched: a) Poor spectra quality with few peaks having low signal-to-noise ratio. These peaks cannot be submitted to MS/MS experiments. b) Good quality spectra, but the PMF and PFF did not return any match to a protein homolog because the homolog differs in its amino-sequence. A variation of one amino-acid in the sequence of the homolog proteins will result in two tryptic peptides of different masses and they will not be matched by MASCOT. c) Post-translational modification of the protein. d) The protein has not been reported yet. We

addressed this limitation by performing de novo sequencing on spectra that did not return any match. Unfortunately, the peptide sequences resulting from de novo sequencing were too short to generate meaningful matches when submitted to BLAST searches. [39, 40]

With large datasets and multiple identifications, there was concern that some of the identifications were incorrect matches due to random chance. To address this concern, we submitted each sample to a target-decoy search to evaluate the false positive identification rate [41]. A target-decoy database is a "randomized" protein database that is created from the original protein database. It has the same number of proteins, which in turn have the same size except that their amino acid sequences have been shuffled. MASCOT offers the option of conducting the search against a concatenated database that contains both original and randomized proteins or using a dedicated decoy protein database. We chose to use the concatenated database, because it would not only take into consideration the score of the protein identification but its ranking against other random and non-random proteins. We considered a match as being a "false positive identification" when MASCOT matched the MS and MS/MS spectra to a randomized protein sequence and ranked it first amongst all possible identifications.

The results of the decoy database search are summarized in table 5.1. The first and main result was that no identified spot matched the decoy database with a higher score than a match against the actual protein sequence. In fact, three of the five protein datasets (proteins originating from muscle, brain, and heart) resulted in no matches to the decoy database. For liver and gill, a single decoy hit was returned with MASCOT scores equal or higher than 67. For gill, this hit (score of 68) was found when searching the first 96 gel spots (more abundant), while for liver (score of 72) the hit was found searching the second 96 gel spots (less abundant). Thus, out of a total of 864 database searches, only two resulted in hits against a randomized protein sequence, corresponding to 0.23% of all searches. If one assumes that

an equal number of random hits occurred against the real sequences, then the actual misidentification rate is expected to be 0.46%, or more than one-order of magnitude less than the commonly accepted P value of 5%.

A second method for evaluating the results of the target-decoy searches has been described by Elias and Gygi [41]. Any protein that was identified from the decoy database was called a "decoy hit" and they assumed that the total number of false positive in a target decoy search should take into consideration false positive matches in the "target database" that are "hidden". The number of false positive will be obtained from doubling the number of decoy hits assuming that incorrect assignments from target and decoy are equally likely to happen. Following the approach of Elias and Gygi, the total number of "False Positive" (FP) will be four. The number of "True Positives" (TP) will be equal to the total assignments (405) minus the false positives (4), or a total of 401. Finally, the False Positive rate will equal FP/TP (4/401 = 0.00987) or just less than 1%. By this measure, it is likely that four out of the 405 identified proteins are incorrect assignments.

Another way to look at these results is to compare the highest MASCOT score for a "decoy hit" (72) with the distribution of MASCOT scores for "real" identifications (fig. 5.2). Of the 405 spots considered identified, 35 had MASCOT scores between 67 and 72. An extremely conservative approach would be to exclude all of these (8.6% of the protein assignments). We feel that this approach is overly conservative and would result in correct assignments that were excluded (i.e., inflating our false negative rate).

5.5 Conclusions

We were successful at implementing a protocol for false discovery rate (FDR) assessment for a proteomic project profiling protein expression in five tissues of the Gulf killifish, *Fundulus grandis*. Three of the five tissues returned no hits against the decoy database, while searches of liver and gill MS data returned one hit per tissue against the decoy database. Using the approach described by Elias and Gygi our False Positive rate was less than 1%, suggesting that more than 99% of the 405 gel spots were correctly identified. This result confirms the view that the vast majority of the protein assignments made by database searching with MASCOT are correct. Nevertheless, those identifications having MASCOT scores equal or less than 72, the highest score for a decoy hit in our experiment, should be viewed with caution and it would be desirable to verify their identities by other approaches (e.g. Western blotting).

5.6 References

[1] Kellner, R. Proteomics. Concepts and perspectives. *Fresenius Journal of Analytical Chemistry* **2000**, 366, 517-524

[2] Tyers, M., Mann, M. From genomics to proteomics. *Nature* **2003**, 422, 193-197

[3] Zhu, H., Bilgin, M., Snyder, M. Proteomics. *Annual Review of Biochemistry.***2003**, 72, 783-812

[4] Van Roessel, P., Brand, A, H. Imaging into the future: visualizing gene expression and protein interactions with fluorescent proteins. *Nature Cell Biology* **2002**, 4, E15

[5] Lander *et al.*; Initial sequencing and analysis of the human genome. *Nature* **2001**, 409, 860-921

[6] Venter *et al*, The sequence of the human genome. *Science* **2001**, 291, 1304-1351

[7] Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray Ionization For Mass-Spectrometry of Large Biomolecules. *Science* **1989**, 246, 64-71

[8] Karas, M.; Hillenkamp, F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10,000 Daltons. *Analytical Chemistry* **1988**, 60, 2299-2301

[9] Domon, B., Aebersold, R.; Mass Spectrometry and Protein Analysis. *Science* **2006**, 312, 212-217

[10] Aerbersold, R., Mann, M. Mass Spectrometry-based proteomics. *Nature* **2003**, 422, 199-207

[11] Gstaiger, M., Aebersold, R. Applying mass spectrometry-based proteomics to genetics, genomics andnetwork biology. *Nature Review Genetics* **2009**, 10, 617-627

[12] Frohlich, T., Arnold, G. J.; Proteome research based on modern liquid-chromatography-tandem mass spectrometry: separation, identification and quantification. *Journal of Neural Transmission* **2006**, 113, 973-994

[13] Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y., Aebersold, R.; Evaluation of two-dimensional gel electrophoresis-based proteome analysis technologoy. *Proceedings of the National Academy of Science* **2000**, 97, 9390-9395

[14] Mc Donalds, W. H., Yates, J. R.; Shotgun Proteomics and Biomarkers Discovery; *Disease Markers* 2002, 18, 99-105

[15] Marcotte, E. M. How do shotgun proteomics algorithms identify proteins? *Nature Biotechnology* **2007**, 25, 755-757

[16] Dworzanski, J. P., Deshpande, S. V., Chen, R., Jabbour, R. E., Snyder, A. P., Wick, C. H., Li, L.; Mass Spectrometry-Based Proteomics Combined with bioinformatics Tools for Bacterial Classification. *Journal of Proteome Research* **2006**, 5, 76-87

[17] Thiede, B., Hohenwater, W., Krah, A., Mattow, J., Schmid, M., Schmidt, F., Jungblut, P. R. Peptide mass fingerprinting. *Methods* **2005**, 35, 237-247

[18] Eng, J. K.,Mc Cormack, A. L., Yates, J. R. An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. *Journal of the American Society for Mass Spectrometry* **1994**, 5, 976-989

[19] Clauser, K. R., baker, P., Burlingame, A. L. Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS MS and database searching. *Analytical Chemistry* **1999**, 71, 2871-2872

[20] Perkins, D. N., Pappin, D. J. C., Creasy, D. M., Cottrell, J.S.; Probability-based protein identification by searching sequence database using mass spectrometry data. *Electrophoresis* **1999**, 20, 3551-3567

[21] Papin, D. J. C., Hojrup, P., Bleasby, A. J.; Rapid identification of proteins by peptide-mass fingerprinting. *Current Biology* 1993, 3, 327-332

[22] Patterson, S. D. Data Analysis – The Achilles heel of proteomics. *Nature Biotechnology* **2003**, 422, 233-237

[23] Carr, S., Aebersold, R., Baldwin, M., Burlingame, A., Clauser, K., Nesvizhskii, A. The need for guidelines in publication of peptide and protein identification data. *Molecular Cell Proteomics* **2004**, 3, 351-353

[24] Tomanek, L. Environmental Proteomics: Changes in the proteome of Marine Organisms in Response to Environmental Stress, Pollutants, Infection, Symbiosis, and Development. *The Annual Review of Marine Science* 2011, 3, 373-399

[25] Wang, J., Wei, Y., Wang, D., Chan, L. L., Dai, J.; Proteomic Study of the effects of complex environmental stresses in the livers of goldfish (Carassius auratus) that inhabit Gaobeidan Lake in Beijing, China. *Ecotoxicology* 2008, 17(3), 213-220

[26] Tomanek, L. Variation of the heat shock response and its implication for predicting the effect of global climate change on species' biogeographical distribution ranges and metabolic costs. *Journal of Experimental Biology* **2010**, 213, 971-979

[27] Kultz, D., Fiol, D., Valkova, N., Gomez-Jimenez, S., Chan, S. Y., Lee, J.; Functional genomics and proteomic of the cellular osmotic stress response in 'non-model' organisms. *The Journal of Experimental Biology*. 2007, 210, 1593-1601

[28] Wulff, T., Jessen, F., Roepstorff, P., Hoffman, E. K.; Long term anoxia in rainbow trout investigated by 2-DE and MS/MS-based proteome approach. *Proteomics* **2008**, 8, 1009-1018

[29] Link, V., Schevchenko, A., Heisenberg, C.; Proteomics of early zebrafish embryos. *BMC Developmental Biology* **2006**, 6, 1-9

[30] Tay, T. L., Lin, Q., Seow, T. K., Tan, K. H. et al., Proteomic analysis of protein profiles during early development of the zebrafish, Danio rerio. *Proteomics* 2006, 6, 3176–3188

[31] Ziv, T., Gattegno,T., Chapovestky, V., Wolf, H., Barnea, E., Lubzens, E., Admon, A.; Comparative proteomics of the developing fish (zebrafish and gilthead seabream) oocytes. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*. 2008, 3(1), 12-35

[32] Burnett, K. et al.; Fundulus as the Premier Teleost Model in Environmental Biology: Opportunities for new Insights Using Genomics. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*. 2007, 2(4), 257-286

[33] Abbaraju, Naga. Patterns of protein expression in tissues of the killifish, *Fundulus heteroclitus* and *Fundulus grandis*. PhD dissertation. Chapter

[34] Laemmli, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 1970, 227, 680-685

[35] Pearson, W. R.; Lipman, D. J.; Improved tools for biological sequence comparison. *Proceedings of the National Academy of Science.* **1988**, 85, 2444-2448

[36] Altschul, S. F.; Madden, T.; Schaffer, A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.; Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*. **1997**, 25(17), 3389-3402

[37] http://www.matrixscience.com/help/decoy_help.html

[38] Pappin, D. J., Hojrup, P., Bleasby, A. J.; Rapid identification of proteins by peptide-mass fingerprinting. *Current Biology* **1993**, 3, 487

[39] Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman. Basic Local Alignment Search Tool. *Journal of Molecular Biology* **1990**, 215, 403-410

[40] Shevchenko, A.; Sunyaev, S.; Loboda, A.; Shevchenko, A.; Bork, P.; Ens, W.; Standing, K. G.; Charting the Proteomes of Organisms with Unsequenced Genomes by MALDI-Quadrupole Time-of-Flight Mass Spectrometry and BLAST Homology searching. *Analytical Chemisrty* **2001**, 73, 1917-1926

[41] Elias, J.; Gygi, S. P.; Target-Decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods* **2007**, 4, 207-214

Appendix

Table 1. List of total identified proteins from MASCOT database search of *Fundulus grandis* liver.

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion score)	Mr (kDa)	pI	Accession No.	Species matched in
1	Beta actin	527	49%	12 (10)	42050	5.29	gi|27805142	*Dicentrarchus labrax*
1	Glucose-regulated protein 94	426	12%	10 (10)	92180	4.7	gi|110226526	*Danio rerio*
1	Triosephosphate isomerase B	408	60%	14(7)	26873	6.9	gi|126211567	*Poecilia reticulata*
1	F1 ATP synthase beta subunit	367	33%	12 (10)	53949	5.15	gi|226441961	*Gillichthys seta*
1	Triosephosphate isomerase B	318	38%	6 (5)	26761	7.6	gi|15149252	*Xiphophorus maculatus*
1	Enolase 1 isoform b	315	27%	7 (7)	43940	5.6	gi|226441957	*Gillichthys seta*
1	Heart-type fatty acid-binding protein	276	40%	6 (5)	14804	5.74	gi|15072477	*Fundulus heteroclitus*
1	Methionine adenosyltransferase I, alpha	268	23%	7 (6)	43262	6.32	gi|41054081	*Danio rerio*
1	Glutathione-S-transferase theta	262	51%	5 (5)	16335	5.7	gi|52001203	*Fundulus heteroclitus*
1	F1 ATP synthase beta subunit	231	37%	13 (5)	53949	5.15	gi|226441961	*Gillichthys seta*
1	Aldolase-B	230	70%	6 (5)	20757	5.96	gi|24473730	*Fundulus heteroclitus*
1	Phosphoenolpyruvate carboxykinase	227	15%	7 (7)	69944	8.79	gi|225579625	*Lateolabrax japonicus*
1	Aldolase-B	218	70%	8 (6)	20757	5.96	gi|24473730	*Fundulus heteroclitus*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
1	79 kDa glucose-regulated protein	202	18%	11 (5)	73981	5.05	gi|47085775	*Danio rerio*
1	Enolase A	215	18%	6 (5)	47469	5.98	gi|98979415	*Danio rerio*
1	Beta-enolase	204	22%	7 (6)	20757	5.96	gi|47210809	*Osmerus mordax*
1	Glyceraldehyde 3-phosphate dehydrogenase	196	32%	8 (4)	36094	7.74	gi|215261537	Kryptolebias marmoratus
1	Tubulin alpha chain	189	24%	7 (3)	50740	4.97	gi|209155464	*Salmo salar*
1	Glycogen phosphorylase, liver form	189	20%	16 (5)	94102	6.56	gi|47221287	*Gallus gallus*
1	Heat shock cognate 70	186	28%	15 (5)	71052	5.27	gi|77999572	*Fundulus heteroclitus macrolepidotus*
1	Ferritin M subunit	184	27%	4 (3)	20642	5.7	gi|4585816	*Larimichthys crocea*
1	Protein disulfide-isomerase	174	11%	6 (4)	55176	4.74	gi|47223959	*Danio rerio*
1	Tubulin alpha-4A chain	163	22%	7 (5)	64743	4.99	gi|223648222	*Salmo salar*
1	Glyceraldehyde 3-phosphate dehydrogenase	159	21%	4 (3)	36094	7.74	gi|215261537	*Kryptolebias marmoratus*
1	Succinate-CoA ligase, GDP-forming beta subunit	156	10%	5 (4)	46906	5.53	gi|189525094	*Danio rerio*
1	Lactate dehydrogenase B	145	15%	4 (3)	36470	7	gi|388146	*Fundulus heteroclitus*
1	Nucleoside diphosphate kinase	144	26%	3 (3)	16666	6.84	gi|197725753	*Epinephelus coioides*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion score)	Mr (kDa)	pI	Accession No.	Species matched in
1	Glyceraldehyde 3-phosphate dehydrogenase	143	24%	5 (4)	36012	8.55	gi|119943230	*Misgurnus anguillicaudatus*
1	Peroxiredoxin 4-like	139	27%	7 (4)	28913	6.41	gi|292613542	*Danio rerio*
1	S-adenosylmethionine synthase isoform type-1	137	26%	6 (3)	43718	7.38	gi|41054081	*Danio rerio*
1	Peroxiredoxin-1	133	26%	5 (2)	21933	5.91	gi|225708082	*Osmerus mordax*
1	Ferritin M subunit	128	17%	3 (2)	20642	5.7	gi|265385728	*Larimichthys crocea*
1	14-3-3 E1 protein	127	62%	6 (4)	14386	5.15	gi|52001207	*Fundulus heteroclitus*
1	Heat shock protein 60	127	26%	12 (3)	61454	5.41	gi|300679902	*Tanichthys albonubes*
1	Enolase 1 isoform b	125	20%	6 (4)	43940	5.6	gi|226441957	*Gillichthys mirabilis*
1	Glutathione peroxidase 1b	125	22%	3 (40)	16132	7.74	gi|47229604	*Danio rerio*
1	4-hydroxyphenylpyruvate dioxygenase	122	27%	11 (3)	45225	6.33	gi|229366896	*Anoplopoma fimbria*
1	Fructose-1,6-bisphosphatase 1b	121	24%	5 (5)	37124	6.9	gi|47085885	*Danio rerio*
1	Betaine homocysteine methyltransferase	111	10%	2 (2)	44641	6.61	gi|56121765	*Danio rerio*
1	Mitochondrial phosphoenolpyruvate carboxykinase	109	21%	13 (2)	69944	8.79	gi|225579625	*Lateolabrax japonicus*
1	Keratin 18	105	7%	3 (3)	49205	5.6	gi|82191536	*Danio rerio*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion score)	Mr (kDa)	pI	Accession No.	Species matched in
1	Phosphoenolpyruvate carboxykinase	102	7%	3 (3)	69944	8.14	gi|185134359	*Lateolabrax japonicus*
1	Tryptophan hydroxylase	99	19%	10 (2)	53141	5.3	gi|261265283	*Oreochromis niloticus*
1	Aldolase-B	96	25%	2 (2)	20757	5.96	gi|24473730	*Fundulus heteroclitus*
1	Tyrosine 3-monooxygenase	96	30%	10 (2)	27375	4.68	gi|47085905	*Danio rerio*
1	Cytochrome b-c1 complex subunit 1, mitochondrial	95	8%	4 (3)	52142	6.28	gi|41387118	*Danio rerio*
1	3-hydroxyanthranilate 34-dioxygenase	93	21%	8 (3)	33267	5.79	gi|47221724	*Ictalurus punctatus*
1	Calreticulin	92	9%	5 (2)	34431	4.46	gi|323650050	*Perca flavescens*
1	PREDICTED: ERC protein 2-like	92	21%	24 (0)	104799	6.64	gi|47229697	*Danio rerio*
1	Similar to 4-hydroxyphenylpyruvate dioxygenase	89	13%	5 (2)	44179	6.21	gi|41054723	*Danio rerio*
1	Heat shock cognate 70	86	24%	14 (1)	71052	5.27	gi|77999572	*Fundulus heteroclitus macrolepidotus*
1	Triose phosphate isomerase B	85	51%	10 (0)	22945	6.14	gi|46849383	*Oryzias latipes*
1	Beta-enolase	82	12%	3 (3)	47423	5.99	gi|47210809	*Danio rerio*
1	Non-neuronal tryptophan hydroxylase 1	78	10%	4 (1)	53141	5.3	gi|261265283	*Oreochromis niloticus*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion score)	Mr (kDa)	pI	Accession No	Species matched in
1	4-hydroxyphenylpyruvate-dioxygenase	77	26%	9 (1)	45225	6.33	gi|229366896	*Anoplopoma fimbria*
1	Intraflagellar transport protein 81 homolog	76	29%	19 (0)	79185	6.43	gi|50539694	*Danio rerio*
1	Similar to 4-hydroxyphenylpyruvate dioxygenase	75	22%	2 (1)	13261	9.43	gi|10121623	*Gillichthys mirabilis*
1	Catechol-O-methyltransferase domain-containing protein 1	75	5%	1 (1)	28822	8.4	gi|209733938	*Salmo salar*
1	Aldolase-B	73	34%	5 (2)	20757	5.96	gi|24473730	*Fundulus heteroclitus*
1	Phenylalanine hydroxylase	73	16%	8 (1)	51322	5.6	gi|41054599	*Danio rerio*
1	Fumarylacetoacetase	72	12%	5 (2)	49876	6.27	gi|225706644	*Osmerus mordax*
1	Amylase-3 protein	71	6%	3 (3)	57527	6.43	gi|48762520	*Danio rerio*
1	WD repeat-containing protein 65	70	18%	22 (0)	139006	6.2	gi|317418584	*Dicentrarchus labrax*
1	Similar to interleukin-1 receptor-associated kinase 4	70	30%	9 (1)	41742	5.66	gi|41055831	*Danio rerio*
1	Tropomyosin1-2	69	25%	7 (1)	29272	4.93	gi|28557124	*Takifugu rubripes*
2	Triosephosphate isomerase B	244	37%	9 (4)	26873	6.9	gi|126211567	*Poecilia reticulata*
2	Heat shock cognate 70	225	15%	10 (6)	71052	5.27	gi|77999572	*Fundulus heteroclitus macrolepidotus*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
2	Liver fatty acid binding protein	183	42%	5 (3)	14241	9.06	gi|52001213	*Fundulus heteroclitus*
2	Aldolase-B	180	31%	7 (3)	20415	5.96	gi|24473730	*Fundulus heteroclitus*
2	Triose phosphate isomerase B	162	25%	5 (3)	26873	6.9	gi|126211567	*Poecilia reticulata*
2	Keratin, type II cytoskeletal 8	151	41%	14 (3)	35739	4.72	gi|47210017	*Danio rerio*
2	Phosphoglycerate kinase 1	147	26%	8 (3)	44106	7.01	gi|47087077	*Danio rerio*
2	ATP synthase subunit alpha, mitochondrial	145	20%	9 (3)	57212	9.06	gi|213512628	*Salmo salar*
2	Beta-actin	140	37%	7 (2)	25116	5.19	gi|239596158	*Danio rerio*
2	Beta globin	131	17%	3 (2)	16202	7.14	gi|22135542	*Paramisgurnus dabryanus*
2	Hemoglobin beta-A chain	123	26%	4 (2)	16079	7.82	gi|67772043	*Siniperca chuatsi*
2	S-adenosylmethionine synthase isoform type-1	122	14%	5 (3)	43001	6.6	gi|47223302	*Danio rerio*
2	Glutathione-S-transferase theta	121	18%	2 (2)	16164	5.7	gi|52001203	*Fundulus heteroclitus*
2	Tubulin beta-2C chain	118	15%	7 (2)	50204	4.76	gi|223672899	*Salmo salar*
2	Alanine-glyoxylate aminotransferase	111	15%	3 (2)	30143	8.93	gi|60417200	*Platichthys flesus*
2	Catalase	108	15%	8 (4)	60144	7.27	gi|225698216	*Rachycentron canadum*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion score)	Mr (kDa)	pI	Accession #	Species matched in
2	Manganese superoxide dismutase	107	25%	5 (2)	25326	8.61	gi|237847752	*Hemibarbus mylodon*
2	Ferritin heavy subunit	95	23%	5 (2)	21102	5.66	gi|115344220	*Epinephelus awoara*
2	Beta actin	93	32%	8 (2)	32193	5.23	gi|164609101	*Squalius alburnoides*
2	Skeletal muscle tropomyosin1-2	92	38%	13 (1)	32783	4.69	gi|22415765	*Takifugu rubripes*
2	Liver fatty acid binding protein	92	50%	6 (3)	14241	9.06	gi|52001213	*Fundulus heteroclitus*
2	Profilin 2	78	7%	2 (2)	15223	7.66	gi|41152213	*Danio rerio*
2	Glutathione-S-transferase theta	78	18%	2 (2)	16335	5.7	gi|52001203	*Fundulus heteroclitus*
2	Aldehyde dehydrogenase family 7 member A1 homolog	77	17%	10 (1)	59455	7.1	gi|213514574	*Salmo salar*
2	Similar to 4-hydroxyphenylpyruvate dioxygenase	75	17%	7 (2)	44005	6.21	gi|41054723	*Danio rerio*
2	Hypothetical protein LOC100126128	72	50%	10 (0)	20380	5.64	gi|157954496	*Danio rerio*
2	Dihydrolipoyl dehydrogenase, mitochondrial precursor	71	15%	8 (2)	54406	7.6	gi|209154114	*Salmo salar*

Note: Proteins highlighted in yellow were identified subsequent to original analysis presented in Abbaraju [Abbaraju, N. (2011) Patterns of protein expression in tissues of the killifish, *Fundulus heteroclitus* and *Fundulus grandis*. Doctoral dissertation, University of New Orleans]

Table 2. List of total identified proteins from MASCOT database search of *Fundulus grandis* gill

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion score)	Mr (kDa)	pI	Accession No	Species matched in
1	Triosephosphate isomerase B	757	52%	12 (8)	26945	6.9	gi|47271422	*Danio rerio*
1	Beta actin	606	50%	17 (7)	42012	5.38	gi|8886013	*Oncorhynchus mykiss*
1	Tubulin beta-1 chain	564	61%	23 (9)	50211	4.79	gi|223672699	*Salmo salar*
1	Skeletal alpha-actin	559	48%	14 (6)	41843	5.28	gi|6653228	*Sparus aurata*
1	Muscle-type creatine kinase	513	44%	18 (7)	43123	6.32	gi|255502901	*Siniperca chuatsi*
1	Heat shock cognate 70	510	52%	26 (7)	71052	5.27	gi|77999572	*Fundulus heteroclitus macrolepidotus*
1	Beta-actin	502	51%	15 (6)	42038	5.29	gi|119943232	*Misgurnus anguillicaudatus*
1	Fructose-bisphosphate aldolase A	422	36%	12 (6)	40128	8.59	gi|317419331	*Dicentrarchus labrax*
1	Triosephosphate isomerase B	354	62%	13 (6)	26761	7.6	gi|15149252	*Xiphophorus maculatus*
1	Krt4 protein	345	24%	13 (5)	53980	5.34	gi|161612220	*Danio rerio*
1	F1 ATP synthase beta subunit	343	51%	17 (5)	53949	5.15	gi|226441961	*Gillichthys seta*
1	Keratin, type II cytoskeletal 8	340	43%	13 (4)	35682	4.72	gi|41056085	*Danio rerio*
1	Krt4 protein	323	29%	16 (6)	53980	5.34	gi|161612220	*Danio rerio*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
1	Skeletal alpha-actin	316	36%	11 (5)	42185	5.28	gi|6653228	*Sparus aurata*
1	Fructose-bisphosphate aldolase C	310	30%	8 (5)	36037	6.36	gi|46849375	*Oryzias latipes*
1	Skeletal muscle actin	306	35%	11 (4)	42185	5.28	gi|6653228	*Sparus aurata*
1	78 kDa glucose-regulated protein	302	39%	24 (5)	74152	5.05	gi|213511032	*Salmo salar*
1	Tropomyosin alpha-3 chain isoform 1	278	37%	13 (5)	32927	4.71	gi|45387763	*Danio rerio*
1	Myosin, light polypeptide 2, skeletal muscle	265	71%	11 (3)	16639	4.39	gi|18859049	*Danio rerio*
1	Malate dehydrogenase, mitochondrial	263	38%	11 (5)	35966	8.56	gi|47085883	*Danio rerio*
1	Tubulin alpha 8-like 3a protein	254	44%	15 (5)	50719	4.93	gi|267567402	*Danio rerio*
1	Slow myotomal muscle tropomyosin	232	25%	12 (5)	32661	4.71	gi|3063940	*Salmo trutta*
1	Tropomyosin2	207	30%	12 (4)	32556	4.7	gi|28557126	*Takifugu rubripes*
1	Glyceraldehyde-3-phosphate dehydrogenase	205	29%	8 (3)	36028	7.81	gi|185133678	*Salmo salar*
1	Keratin, type II cytoskeletal 8	205	36%	11 (2)	35739	4.72	gi|41056085	*Danio rerio*
1	Fructose-bisphosphate aldolase A	203	25%	13 (4)	39742	8.45	gi|41282154	*Danio rerio*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
1	Warm temperature acclimation-related 65kDa protein	195	17%	6 (3)	48298	5.35	gi|86610887	*Xiphophorus hellerii*
1	Tropomyosin4-1	194	38%	13 (3)	28681	4.65	gi|28557136	*Takifugu rubripes*
1	Peroxiredoxin-1	192	27%	6 (3)	22019	6.51	gi|229366432	*Anoplopoma fimbria*
1	Lactate dehydrogenase B	171	44%	11 (4)	36486	7	gi|388144	*Fundulus heteroclitus*
1	Keratin 18	159	25%	9 (2)	33547	4.75	gi|189498197	*Epinephelus coioides*
1	Keratin, type II cytoskeletal 8	145	33%	10 (2)	35739	4.72	gi|41056085	*Danio rerio*
1	Putative cytosolic carbonic anhydrase	143	36%	4 (2)	8064	9.3	gi|302594566	*Cyprinodon variegatus*
1	Muscle-type creatine kinase CKM2	142	30%	11 (3)	42980	6.44	gi|21694043	*Oreochromis mossambicus*
1	Nuclease diphosphate kinase B	135	33%	4 (2)	17214	6.82	gi|10121713	*Gillichthys mirabilis*
1	Type I cytokeratin, enveloping layer, like	127	15%	4 (1)	41617	5.17	gi|130504059	*Danio rerio*
1	Beta-actin	124	24%	7 (2)	42012	5.38	gi|8886013	*Oncorhynchus mykiss*
1	Coronin-1A	111	15%	7 (2)	51371	6.24	gi|41055464	*Danio rerio*
1	Low molecular weight heat shock protein Hsp27	107	12%	3 (2)	22634	6.24	gi|1835583	*Poeciliopsis lucida*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
1	Heterogeneous nuclear ribonucleoprotein A/B	106	19%	8 (2)	35331	5.67	gi|108744009	*Haplochromis burtoni*
1	WD repeat-containing protein 1	104	25%	12 (1)	67272	5.93	gi|317419065	*Dicentrarchus labrax*
1	Glyceraldehyde-3-phosphate dehydrogenase	100	25%	7 (2)	36028	7.81	gi|185133678	*Salmo salar*
1	Phosphoglycerate kinase 1	99	20%	7 (2)	44505	7.01	gi|47087077	*Danio rerio*
1	Cytoplasmic carbonic anhydrase	98	5%	1 (1)	28391	5.46	gi|41059441	*Oncorhynchus mykiss*
1	Full=Calmodulin-A; Short=CaM A	85	33%	4 (2)	15348	4.08	gi|49037466	*Oryzias latipes*
1	Proteasome subunit alpha type-4	81	44%	9 (1)	29599	7.57	gi|47550827	*Anoplopoma fimbria*
1	Adenylyl cyclase-associated protein 1	76	26%	9 (1)	49707	8.52	gi|41054003	*Danio rerio*
1	Serine/cysteine proteinase inhibitor	71	14%	3 (1)	22858	5.01	gi|52430384	*Fundulus heteroclitus*
1	Myosin, heavy polypeptide 10, non-muscle	68	12%	20 (1)	268737	5.68	gi|292614613	*Fundulus heteroclitus*
1	COMM domain-containing protein 1	68	61%	8 (0)	21334	5.48	gi|113674991	*Danio rerio*
2	Receptor for activated protein kinase C	465	83%	21 (7)	35541	8.07	gi|3023850	*Oreochromis niloticus*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
2	Beta actin	286	38%	10 (4)	37363	5.46	gi|45237481	*Hippoglossus hippoglossus*
2	F-actin-capping protein subunit beta	265	47%	13 (5)	30766	5.5	gi|229366390	*Anoplopoma fimbria*
2	Adenylate kinase	205	34%	7 (4)	21335	6.19	gi|225707436	*Osmerus mordax*
2	Enolase 1 isoform b	182	45%	12 (3)	43540	5.9	gi|226441955	*Gillichthys mirabilis*
2	Proteasome subunit alpha type 2	172	36%	6 (3)	25940	6.12	gi|225706328	*Osmerus mordax*
2	Fructose-bisphosphate aldolase A	172	34%	12 (2)	39631	8.5	gi|295792244	*Epinephelus coioides*
2	Proteasome subunit beta type-1-A	145	52%	11 (3)	26443	6.52	gi|17380207	*Carassius auratus*
2	Beta-enolase	135	35%	12 (4)	47423	5.99	gi|47551317	*Danio rerio*
2	Pyruvate kinase muscle isozyme	133	30%	13 (3)	57436	5.97	gi|224587654	*Salmo salar*
2	Nucleoside diphosphate kinase	123	26%	3 (3)	16551	6.84	gi|197725753	*Epinephelus coioides*
2	Type II basic cytokeratin	118	20%	9 (2)	54242	5.42	gi|18858947	*Danio rerio*
2	Proteasome subunit alpha type-6	95	38%	8 (2)	27375	6.45	gi|229367144	*Anoplopoma fimbria*
2	Stress-induced phosphoprotein 1	79	19%	12 (3)	62248	6.49	gi|310693634	*Miichthys miiuy*
2	TBT-binding protein	71	18%	2 (1)	19702	6.32	gi|52001215	Fundulus heteroclitus

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
2	Actin, beta-like 2	69	30%	8 (1)	41979	5.18	gi|50344802	*Danio rerio*
2	Hypothetical protein LOC100124602	68	20%	7 (1)	28792	4.83	gi|156616350	*Danio rerio*
2	Type II keratin E3-like protein	68	19%	6 (1)	38600	4.89	gi|48476437	*Sparus aurata*
2	Ran-specific GTPase-activating protein	67	25%	5 (1)	27404	5.74	gi|47086517	*Danio rerio*
2	Upf0553 protein c9orf64-like protein	67	32%	11 (0)	39512	5.36	gi|318209996	*Ictalurus punctatus*

Note: Proteins highlighted in yellow were identified subsequent to original analysis presented in Abbaraju'[Abbaraju, N. (2011) Patterns of protein expression in tissues of the killifish, *Fundulus heteroclitus* and *Fundulus grandis*. Doctoral dissertation, University of New Orleans]

Table 3. List of total identified proteins from MASCOT database search of *Fundulus grandis* heart.

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mw (kDa)	pI	Accession No.	Species matched in
1	Fast myotomal muscle actin	693	70%	18 (9)	42290	5.22	gi|10953948	*Salmo salar*
1	F1 ATP synthase beta subunit	667	64%	23 (9)	55223	5.09	gi|226441961	*Gillichthys seta*
1	Triosephosphate isomerase B	583	65%	13 (8)	26945	6.9	gi|47271422	*Danio rerio*
1	Acta1a protein	517	64%	13 (7)	41890	5.29	gi|182890476	*Danio rerio*
1	Isocitrate dehydrogenase NADP, mitochondrial	470	37%	18 (8)	46823	6.57	gi|47224185	*Danio rerio*
1	Lactate dehydrogenase B	461	58%	15 (8)	36442	7.64	gi|388130	*Fundulus heteroclitus*
1	Skeletal alpha-actin	417	49%	16 (6)	41843	5.28	gi|6653228	*Sparus aurata*
1	Aldolase A fructose-bisphosphate	411	26%	8 (2)	40228	8.45	gi|37595414	*Danio rerio*
1	Creatine kinase	359	36%	15 (8)	43317	5.98	gi|124358786	*Fundulus heteroclitus*
1	Novel protein similar to vertebrate tropomyosin 4 (TPM4) (zgc:63909)	331	48%	13 (5)	23518	4.6	gi|94732584	*Danio rerio*
1	Actin alpha sarcomeric/cardiac	320	51%	16 (4)	41962	5.22	gi|317725965	*Ictalurus punctatus*
1	Enolase 1, alpha	302	43%	14 (5)	47386	6.16	gi|37590349	*Danio rerio*
1	Heat shock cognate 71	289	25%	16 (7)	70710	5.23	gi|77999572	*Fundulus heteroclitus macrolepidotus*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
1	Heat shock protein hsp70	281	48%	14 (7)	71552	5.31	gi|126211563	*Poecilia reticulata*
1	Beta-enolase	278	49%	17 (6)	47423	5.99	gi|47551317	*Danio rerio*
1	Glyceraldehyde-3-phosphate dehydrogenase	251	42%	10 (4)	36067	8.63	gi|21955965	*Anguilla japonica*
1	Isocitrate dehydrogenase NADP, mitochondrial	292	32%	16 (6)	46823	6.57	gi|41054651	*Danio rerio*
1	Enolase 1, alpha	248	46%	16 (5)	47386	6.16	gi|37590349	*Danio rerio*
1	TPA: desmin	230	39%	22 (3)	51467	5.1	gi|33186832	*Takifugu rubripes*
1	Mitochondrial ATP synthase alpha-subunit	218	17%	9 (4)	59528	9.33	gi|14009437	*Cyprinus carpio*
1	TPA: desmin	216	44%	25 (4)	51582	5.1	gi|33186832	*Takifugu rubripes*
1	Unnamed protein producmalate dehydrogenase, mitochondrial	212	35%	11 (5)	35966	8.56	gi|47085883	*Danio rerio*
1	Myosin light chain 6B	208	20%	5 (2)	22912	4.58	gi|148233450	*Danio rerio*
1	Glyceraldehyde-3-phosphate dehydrogenase	204	23%	7 (4)	36168	6.4	gi|51895785	*Astatotilapia burtoni*
1	Fructose-bisphosphate aldolase A	204	29%	11 (3)	39631	8.5	gi|295792244	*Epinephelus coioides*
1	Fructose-bisphosphate aldolase C	201	41%	13 (4)	36037	6.36	gi|46849375	*Oryzias latipes*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
1	Pyruvate kinase muscle isozyme	194	30%	15 (4)	57436	5.97	gi|224587654	*Salmo salar*
1	Warm temperature acclimation-related 65kDa protein	189	17%	6 (3)	477285	5.35	gi|86610887	*Xiphophorus hellerii*
1	ATP synthase subunit alpha, mitochondrial	187	34%	17 (5)	59819	9.23	gi|116325975	*Danio rerio*
1	TPA: desmin	175	45%	20 (4)	51582	5.1	gi|33186832	*Takifugu rubripes*
1	Phosphoglycerate kinase 1	164	20%	7 (2)	44106	7.01	gi|47087077	*Danio rerio*
1	Isocitrate dehydrogenase, mitochondrial precursor	162	34%	11 (5)	50456	7.12	gi|225705994	*Osmerus mordax*
1	Glyceraldehyde 3-phosphate dehydrogenase	160	43%	9 (3)	35914	7.81	gi|89143257	*Salmo salar*
1	Sarcomeric mitochondrial creatine kinase-like	160	21%	4 (3)	20646	9.01	gi|292617951	*Danio rerio*
1	Warm temperature acclimation-related 65kDa protein	142	13%	5 (3)	50066	5.47	gi|86610887	*Xiphophorus hellerii*
1	Aconitate hydratase, mitochondrial	140	15%	12 (4)	87500	6.61	gi|38707983	*Danio rerio*
1	Glyceraldehyde-3-phosphate dehydrogenase	131	32%	7 (4)	36168	6.4	gi|51895785	*Astatotilapia burtoni*
1	Chaperonin Cpn60	126	45%	10 (3)	33309	9.13	gi|21805770	*Danio rerio*
1	Skeletal alpha-actin type-2b	119	34%	10 (3)	42286	5.23	gi|30268609	*Coryphaenoides yaquinae*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
1	Medium-chain specific acyl-CoA dehydrogenase, mitochondrial precursor	119	14%	8 (3)	46030	7.52	gi|223648080	*Salmo sala*
1	Glutamate dehydrogenase 3	109	32%	13 (2)	59956	8.26	gi|21666614	*Oncorhynchus mykiss*
1	Methylmalonate-semialdehyde dehydrogenase	107	20%	9 (3)	56800	6.48	gi|209155966	*Salmo salar*
1	Warm temperature-acclimated 65kDa protein	107	59%	5 (3)	7875	7.08	gi|52430344	*Fundulus heteroclitus*
1	Cytochrome b-c1 complex subunit 1, mitochondrial	106	10%	5 (4)	52712	6.28	gi|18496665	*Dicentrarchus labrax*
1	Putative transferrin	102	19%	14 (2)	76152	6.38	gi|34329603	*Acanthopagrus schlegelii*
1	Aconitate hydratase, mitochondrial	101	11%	9 (3)	84729	6.33	gi|317419173	*Dicentrarchus labrax*
1	Voltage-dependent anion channel 2-2	92	18%	5 (3)	30099	8.85	gi|197632613	*Salmo salar*
1	Pyruvate dehydrogenase E1 component subunit beta, mitochondrial precursor	91	18%	6 (2)	39398	5.52	gi|225715630	*Esox Lucius*
1	Enolase 1, alpha	90	31%	10 (3)	47386	6.16	gi|37590349	*Danio rerio*
1	Electron transfer flavoprotein subunit alpha, mitochondrial precursor	86	15%	4 (2)	35213	6.9	gi|229366746	*Anoplopoma fimbria*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
1	Acetyl-Coenzyme A acyltransferase	84	22%	7 (1)	42160	7.64	gi|225794853	*Perca flavescens*
1	Cytochrome b-c1 complex subunit 1, mitochondrial	83	18%	8 (3)	52712	6.28	gi|41387118	*Danio rerio*
1	Pyruvate kinase muscle isozyme	81	28%	12 (1)	57436	5.97	gi|224587654	*Salmo salar*
1	TPA: desmin	78	36%	21 (1)	51582	5.1	gi|33186832	*Takifugu rubripes*
1	Fructose-bisphosphate aldolase C	78	22%	6 (3)	36037	6.36	gi|46849375	*Oryzias latipes*
1	Superoxide dismutase, mitochondrial precursor	77	15%	4 (2)	24933	7.82	gi|223647010	*Salmo salar*
1	Transferrin variant D	79	11%	8 (1)	73138	5.77	gi|189473161	*Cyprinus carpio*
1	Enoyl-CoA hydratase, mitochondrial	76	17%	4 (2)	36840	5.85	gi|213511865	*Salmo salar*
1	Pyruvate dehydrogenase E1 alpha 1	72	20%	11 (2)	43265	7.56	gi|53749653	*Danio rerio*
1	Putative dystrophin	72	19%	50 (1)	419844	5.41	gi|30315803	*Takifugu rubripes*
1	Peroxiredoxin-1	70	19%	4 (2)	21791	6.51	gi|229366432	*Anoplopoma fimbria*
1	Transferrin variant A	69	23%	14 (1)	75595	5.75	gi|18034630	*Cyprinus carpio*
1	Fumarate hydratase, mitochondrial precursor	68	19%	6 (2)	54997	8.98	gi|41055718	*Danio rerio*
1	Apolipoprotein A-I	68	14%	4 (2)	29082	5.29	gi|193795856	*Oplegnathus fasciatus*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
2	ATP synthase subunit alpha, mitochondrial	306	27%	12 (5)	59707	9.08	gi|116325975	*Danio rerio*
2	Prohibitin-like	263	35%	9 (6)	29667	5.28	gi|41351079	*Danio rerio*
2	Triosephosphate isomerase B	248	36%	7 (5)	26588	6.9	gi|126211567	*Poecilia reticulata*
2	NADH dehydrogenase [ubiquinone] iron-sulfur protein 3	194	27%	7 (4)	29380	5.92	gi|62955483	*Danio rerio*
2	Adenylate kinase	181	27%	6 (4)	21407	7.66	gi|213512310	*Salmo salar*
2	Beta tubulin	164	24%	8 (5)	45635	5.35	gi|226441963	*Gillichthys seta*
2	Aconitate hydratase, mitochondrial	150	12%	11 (4)	86872	6.61	gi|38707983	*Danio rerio*
2	Heat shock cognate 70	131	16%	9 (3)	70824	5.27	gi|77999572	*Fundulus heteroclitus macrolepidotus*
2	C-type lectin	130	24%	5 (2)	23919	5.41	gi|52430374	*Fundulus heteroclitus*
2	Glyceraldehyde 3-phosphate dehydrogenase	118	34%	8 (2)	35923	7.74	gi|215261537	*Kryptolebias marmoratus*
2	Skeletal alpha-actin	113	27%	8 (2)	41843	5.28	gi|6653228	*Sparus aurata*
2	Keratin	112	22%	12 (64)	61228	5.04	gi|59710087	*Takifugu rubripe*
2	Novel protein similar to vertebrate tropomyosin 4 (TPM4) (zgc:63909)	111	32%	10 (4)	23518	4.6	gi|94732584	*Danio rerio*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
2	Glial fibrillary acidic protein; GFAP	111	35%	7 (2)	24887	4.9	gi|435737	*Cyprinus carpio*
2	Glutamate dehydrogenase 3	104	19%	10 (4)	59557	8.26	gi|21666614	*Oncorhynchus mykiss*
2	Heat shock cognate 70	102	23%	13 (2)	70824	5.27	gi|77999572	*Fundulus heteroclitus macrolepidotu*
2	Phosphoglycerate mutase 2 muscle	100	27%	5 (3)	28806	8.83	gi|41056123	*Danio rerio*
2	Fructose-bisphosphate aldolase A	92	26%	9 (2)	39229	8.5	gi|295792244	*Epinephelus coioides*
2	Mitochondrial aspartate aminotransferase	91	17%	9 (2)	47429	9.18	gi|308321187	*Ictalurus furcatus*
2	Similar to isocitrate dehydrogenase 3 NAD+ alpha	82	27%	7 (2)	39960	7.04	gi|29124437	*Danio rerio*
2	Heterogeneous nuclear ribonucleoprotein A/B isoform 1	76	11%	5 (2)	34792	5.21	gi|146260280	*Mus musculus*
2	Mitochondrial ATP synthase alpha subunit	72	11%	3 (2)	39511	9.18	gi|44969408	*Danio rerio*
2	Mitochondrial ATP synthase alpha-subunit	69	10%	6 (2)	59528	9.33	gi|14009437	*Cyprinus carpio*
2	C-type lectin	69	21%	5 (1)	23919	5.41	gi|52430374	*Fundulus heteroclitus*
2	Four and a half LIM domains protein 2	69	8%	3 (2)	32590	8.36	gi|229365762	Anoplopoma fimbria

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
2	NADH-ubiquinone oxidoreductase 75 kDa subunit, mitochondrial precursor	68	11%	7 (3)	80139	5.66	gi|209155396	*Salmo salar*
2	Cofilin-2	67	31%	4 (1)	18903	6.18	gi|225706552	*Osmerus mordax*

Note: Proteins highlighted in yellow were identified subsequent to original analysis presented in Abbaraju [Abbaraju, N. (2011) Patterns of protein expression in tissues of the killifish, *Fundulus heteroclitus* and *Fundulus grandis*. Doctoral dissertation, University of New Orleans]

Table 4. List of total identified proteins from MASCOT database search of *Fundulus grandis* brain.

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
1	Beta tubulin	508	49%	19 (9)	49757	4.78	gi|10242162	*Notothenia coriiceps*
1	Heat shock cognate 70	463	45%	23 (7)	70824	5.27	gi|77999572	*Fundulus heteroclitus macrolepidotus*
1	Beta actin	411	48%	14 (7)	41708	5.29	gi|27805142	*Dicentrarchus labrax*
1	Triosephosphate isomerase B	384	44%	10 (6)	26476	7.6	gi|15149252	*Xiphophorus maculatus*
1	ATP synthase subunit alpha, mitochondrial	296	28%	12 (5)	59648	9.23	gi|116325975	*Danio rerio*
1	Triosephosphate isomerase B	294	39%	9 (5)	26476	7.6	gi|15149252	*Xiphophorus maculatus*
1	Beta tubulin	285	31%	12 (6)	49718	4.74	gi|10242186	*Chionodraco rastrospinosus*
1	V-ATPase subunit A	284	35%	19 (7)	68410	5.43	gi|14915706	*Fundulus heteroclitus*
1	Voltage-dependent anion-selective channel protein 1	272	35%	8 (5)	30538	6.53	gi|47777306	*Danio rerio*
1	Enolase 1 isoform a	253	45%	14 (4)	43511	5.16	gi|226441951	*Gillichthys mirabilis*
1	F1 ATP synthase beta subunit	248	46%	17 (4)	53949	5.15	gi|226441961	*Gillichthys seta*
1	Brain creatine kinase b	227	20%	8 (4)	42723	5.39	gi|238231559	Oncorhynchus mykiss
1	Malate dehydrogenase, mitochondrial	225	30%	9 (5)	35510	8.56	gi|47085883	*Danio rerio*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
1	Gamma-enolase	224	19%	7 (6)	56408	5.42	gi|213513750	*Salmo salar*
1	Voltage-dependent anion-selective channel protein 1	223	32%	7 (5)	30538	6.53	gi|47221743	*Danio rerio*
1	Cofilin 2, like	220	38%	5 (4)	18742	6.82	gi|47271384	*Danio rerio*
1	Beta-actin	219	39%	11 (5)	41727	5.16	gi|18034011	*Morulius calbasu*
1	Aconitate hydratase, mitochondrial	218	17%	13 (3)	84729	6.33	gi|317419173	*Dicentrarchus labrax*
1	Pyruvate kinase muscle isozyme	217	28%	15 (5)	57436	5.97	gi|224587654	*Salmo salar*
1	Lactate dehydrogenase B	209	38%	10 (4)	36157	7.64	gi|388130	*mummichog*
1	Enolase 1 isoform a	209	35%	10 (4)	43511	5.16	gi|226441951	*Gillichthys mirabilis*
1	Creatine kinase B-type	201	13%	5 (4)	42912	5.4	gi|225706018	*Osmerus mordax*
1	Creatine kinase	193	23%	11 (4)	46453	8.73	gi|8575804	*Takifugu rubripes*
1	V-type ATPase B subunit	180	30%	13 (5)	55703	5.46	gi|4929105	*Oncorhynchus mykiss*
1	Cofilin 2, like	173	33%	5 (4)	18742	6.82	gi|47271384	*Danio rerio*
1	Adenylate kinase	168	32%	6 (4)	21407	7.66	gi|213512310	*Salmo salar*
1	Muscle cofilin 2	166	19%	3 (2)	18771	6.32	gi|225716038	*Esox lucius*
1	Beta tubulin	164	27%	10 (4)	49757	4.78	gi|10242162	*Notothenia coriiceps*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
1	Gfap protein	164	28%	15 (3)	51087	5.34	gi|46329793	*Danio rerio*
1	Fascin	163	26%	10 (4)	47037	8.35	gi|115494998	*Danio rerio*
1	Dihydropyrimidinase-related protein 3	160	22%	9 (3)	62189	5.91	gi|213511906	*Salmo salar*
1	V-type proton ATPase subunit E 1	156	37%	8 (3)	25974	8.47	gi|27545261	*Danio rerio*
1	Glyceraldehyde-3-phosphate dehydrogenase	151	22%	5 (2)	35826	6.4	gi|51895785	*Astatotilapia burtoni*
1	Creatine kinase, brain	150	20%	6 (4)	42856	5.49	gi|27545193	*Danio rerio*
1	Phosphoglycerate kinase 1	148	23%	9 (2)	44690	6.47	gi|47087077	*Danio rerio*
1	Glutamate dehydrogenase 3	146	27%	13 (3)	59557	8.26	gi|21666614	*Oncorhynchus mykiss*
1	Nucleoside diphosphate kinase	144	26%	3 (4)	16551	6.84	gi|197725753	*Epinephelus coioides*
1	Thioredoxin-dependent peroxide reductase, mitochondrial precursor	143	16%	4 (3)	27147	7.17	gi|225708348	*Osmerus mordax*
1	Triosephosphate isomerase A	143	25%	7 (3)	26748	4.61	gi|15149254	*Xiphophorus maculatus*
1	Enolase 2	138	22%	9 (4)	46812	4.77	gi|51467931	*Danio rerio*
1	Malate dehydrogenase A	137	27%	3 (3)	16650	6.06	gi|62870693	*Oreochromis mossambicus*
1	Tubulin, beta 2	133	28%	11 (4)	50534	4.76	gi|66773102	*Danio rerio*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
1	Peroxiredoxin-1	131	25%	4 (2)	21762	5.91	gi|225708082	*Osmerus mordax*
1	Gefiltin	130	23%	12 (2)	54391	5.23	gi|18858755	*Danio rerio*
1	Alpha tubulin	125	32%	6 (3)	31833	5.35	gi|16517095	*Gillichthys mirabilis*
1	Mitochondrial aspartate aminotransferase	124	27%	11 (3)	47429	9.18	gi|308321187	*Ictalurus furcatus*
1	Glyceraldehyde-3-phosphate dehydrogenase	120	11%	3 (2)	35970	6.2	gi|194241596	*Oplegnathus fasciatus*
1	Malate dehydrogenase, mitochondrial precursor	115	27%	7 (3)	35336	8.36	gi|225708356	*Osmerus mordax*
1	60 kDa heat shock protein, mitochondrial precursor	112	24%	11 (3)	60795	5.56	gi|209153200	*Salmo salar*
1	Fructose-bisphosphate aldolase C	98	16%	6 (3)	35695	6.36	gi|46849375	*Oryzias latipes*
1	Voltage-dependent anion channel	96	11%	1 (1)	18997	8.54	gi|16517086	*Gillichthys mirabilis*
1	Glyceraldehyde-3-phosphate dehydrogenase	95	21%	5 (1)	35826	6.4	gi|51895785	*Astatotilapia burtoni*
1	Gefiltin	94	21%	13 (2)	54391	5.23	gi|18858755	*Danio rerio*
1	Stress-70 protein, mitochondrial	90	21%	10 (2)	51769	5.39	gi|54262125	*Danio rerio*
1	Fascin	89	22%	8 (4)	47037	8.35	gi|115494998	*Danio rerio*
1	Class III intermediate filament protein Danio rerio	87	23%	13 (1)	51255	5028	gi|38374177	*Chaenocephalus aceratus*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
1	Enoyl-CoA hydratase, mitochondrial	86	9%	2 (1)	36327	5.85	gi|213511865	*Salmo salar*
1	Pkm2 protein	86	24%	10 (3)	57436	5.97	gi|224587654	*Salmo salar*
1	Phosphoglycerate kinase	85	15%	5 (1)	44106	7.01	gi|213511822	*Salmo salar*
1	Glial fibrillary acidic protein	80	31%	14 (1)	42127	5.03	gi|40538766	*Danio rerio*
1	Predicted protein	76	7%	3 (1)	95790	7.17	gi|156382697	*Nematostella vectensis*
1	Internexin neuronal intermediate filament protein, alpha	75	15%	10 (1)	64712	4.79	gi|41054736	*Danio rerio*
1	Glutamine synthetase	74	8%	2 (2)	20251	5.33	gi|261363572	*Trachinotus blochii*
1	manganese-containing superoxide dismutase precursor	69	13%	3 (2)	24993	8.29	gi|41152470	*Danio rerio*
1	Calretinin	67	13%	3 (2)	31002	4.8	gi|308323877	*Ictalurus punctatus*
2	Heat shock cognate 70	412	47%	24 (7)	70824	5.27	gi|77999572|	*Fundulus heteroclitus macrolepidotus*
2	ATP synthase subunit alpha, mitochondrial	349	35%	14 (7)	58100	9.3	gi|116325975	*Danio rerio*
2	Tubulin, alpha 8 like 4	313	40%	12 (6)	47867	5.14	gi|41055710	*Danio rerio*
2	Prohibitin	219	39%	8 (5)	29667	5.28	gi|292611600	*Danio rerio*
2	Isocitrate dehydrogenase [NADP], mitochondrial	205	30%	14 (6)	46823	6.57	gi|41054651	*Danio rerio*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
2	Isocitrate dehydrogenase [NADP], mitochondrial	187	29%	13 (5)	46823	6.57	gi|41054651	*Danio rerio*
2	Glyceraldehyde-3-phosphate dehydrogenase	186	14%	4 (3)	35826	6.4	gi|51895785	*Astatotilapia burtoni*
2	Enolase 1 isoform b	178	28%	13 (3)	43540	5.9	gi|226441955	*Gillichthys mirabilis*
2	Brain lipid binding protein	141	36%	5 (3)	14939	5.23	gi|261363562	*Trachinotus blochii*
2	Dihydropyrimidinase-like 5	141	14%	9 (3)	61141	6.17	gi| 223648356	*Salmo salar*
2	Synaptosome-associated protein 25a	138	34%	6 (4)	22843	4.57	gi|56207906	*Danio rerio*
2	Pyruvate dehydrogenase E1 alpha 1	127	17%	10 (3)	43265	7.56	gi|50539866	*Danio rerio*
2	Proteasome subunit alpha type 2	125	34%	6 (3)	25883	6.12	gi|225706328	*Osmerus mordax*
2	B-actin	123	33%	9 (1)	41813	5.3	gi|6693629	*Pagrus major*
2	Fascin	122	12%	6 (5)	47037	8.35	gi|115494998	*Danio rerio*
2	HSP70	121	26%	9 (4)	38979	4.79	gi|126211563	*Poecilia reticulata*
2	Similar to isocitrate dehydrogenase 3 (NAD+) alpha	121	29%	8 (3)	39960	7.04	gi|29124437	*Danio rerio*
2	Beta tubulin	119	20%	8 (4)	49757	4.78	gi|10242162	*Notothenia coriiceps*
2	Pyruvate kinase muscle isozyme	117	23%	12 (2)	57436	5.97	gi|22458765	*Salmo salar*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
2	Heat shock cognate 71	116	43%	7 (2)	21906	7.96	gi|157965668	*Cichlasoma dimerus*
2	Alpha tubulin	112	38%	8 (4)	31833	5.35	gi|16517095	*Gillichthys mirabilis*
2	Voltage-dependent anion channel 2-2	103	10%	5 (3)	30099	8.85	gi|197632613	*Salmo salar*
2	NSFL1 cofactor p47	97	12%	6 (2)	40490	5.04	gi|71894957	*Gallus gallus*
2	Malate dehydrogenase A	96	34%	4 (2)	16650	6.06	gi|62870693	*Oreochromis mossambicus*
2	Proteasome subunit alpha type 1	89	34%	5 (2)	22346	9.08	gi|62079624	*Oreochromis mossambicus*
2	Valosin containing protein	83	21%	15 (2)	89368	5.14	gi|41393119	*Danio rerio*
2	WD repeat-containing protein 1	82	23%	13 (1)	66645	5.93	gi|317419065	*Dicentrarchus labrax*
2	3-monooxygenase/tryptophan 5-monooxygenase activation protein, gamma polypeptide	81	29%	6 (3)	28217	4.86	gi|49227280	*Danio rerio*
2	Mitochondrial ATP synthase alpha-subunit	80	21%	10 (1)	59528	9.33	gi|14009437	*Cyprinus carpio*
2	Dj-1 protein	77	16%	3 (3)	19943	5.83	gi|157278183	*Oryzias latipes*
2	Enolase 1 isoform b	74	18%	7 (1)	43540	5.9	gi|226441955	*Gillichthys mirabilis*
2	Succinyl-CoA ligase subunit alpha, mitochondrial precursor	73	28%	7 (1)	34259	9.43	gi|225707744	*Osmerus mordax*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
2	Dihydropyrimidinase-related protein 3	71	19%	7 (2)	62189	5.91	gi|209155622	*Salmo salar*
2	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydroge	68	10%	4 (2)	44084	6.42	gi|1117958	*Takifugu rubripes*
2	Phosphoglycerate mutase 1	67	19%	4 (1)	29022	6.2	gi|209147645	*Salmo salar*

Note: Proteins highlighted in yellow were identified subsequent to original analysis presented in Abbaraju [Abbaraju, N. (2011) Patterns of protein expression in tissues of the killifish, *Fundulus heteroclitus* and *Fundulus grandis*. Doctoral dissertation, University of New Orleans]

Table 5. List of total identified proteins from MASCOT database search of *Fundulus grandis* muscle.

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
1	Skeletal alpha-actin	718	63%	19 (8)	42185	5.28	gi|6653228	*Sparus aurata*
1	Muscle-type creatine kinase CKM2	634	50%	20 (8)	42980	6.44	gi|21694043	*Oreochromis mossambicus*
1	Muscle-type creatine kinase CKM2	608	46%	19 (7)	42980	6.44	gi|21694043	*Oreochromis mossambicus*
1	Muscle-type creatine kinase CKM2	577	47%	18 (6)	42980	6.44	gi|21694043	*Oreochromis mossambicus*
1	Skeletal alpha-actin	539	53%	15 (7)	41843	5.28	gi|6653228	*Sparus aurata*
1	Skeletal alpha-actin	522	59%	16 (7)	42185	5.28	gi|6653228	*Sparus aurata*
1	Muscle-type creatine kinase CKM2	502	41%	18 (8)	42980	6.44	gi|21694043	*Oreochromis mossambicus*
1	Tropomyosin	487	59%	23 (7)	32767	4.69	gi|60390740	*Pennahia argentata*
1	Triosephosphate isomerase B	477	47%	11 (7)	26945	6.9	gi|47206738	*Danio rerio*
1	Skeletal alpha-actin	460	41%	12 (7)	41843	5.28	gi|6653228	*Sparus aurata*
1	Beta-enolase	453	48%	14 (9)	47423	5.99	gi|47210809	*Danio rerio*
1	Myosin, light polypeptide 2, skeletal muscle	453	69%	12 (7)	16639	4.39	gi|47217809	*Danio rerio*
1	Skeletal alpha-actin	436	55%	16 (7)	42185	5.28	gi|6653228	*Sparus aurata*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
1	Muscle-type creatine kinase CKM2	425	38%	17 (8)	42980	6.44	gi|21694043	*Oreochromis mossambicus*
1	Muscle-type creatine kinase CKM2	378	43%	16 (8)	42980	6.44	gi|21694043	*Oreochromis mossambicus*
1	Enolase 1 isoform b	377	44%	12 (7)	43910	5.16	gi|226441955	*Gillichthys mirabilis*
1	Fructose-bisphosphate aldolase A	357	29%	13 (5)	39742	8.45	gi|41282154	*Danio rerio*
1	Isocitrate dehydrogenase 2 (NADP+), mitochondrial	345	33%	15 (7)	46823	6.57	gi|47224185	*Danio rerio*
1	Muscle-type creatine kinase CKM2	311	41%	16 (5)	42980	6.44	gi|21694043	*Oreochromis mossambicus*
1	myosin light chain 1, skeletal muscle isoform-like	309	35%	7 (4)	20054	4.54	gi|348515637	*Oreochromis niloticus*
1	Phosphoglycerate mutase 2	298	35%	7 (6)	28977	8.83	gi|41056123	*Danio rerio*
1	Skeletal alpha-actin	277	35%	11 (8)	41843	5.28	gi|6653228	*Sparus aurata*
1	Adenylate kinase	267	40%	8 (6)	21521	7.66	gi|213511412	*Salmo salar*
1	Myosin, light polypeptide 2, skeletal muscle	256	63%	10 (5)	16639	4.39	gi|47217809	*Danio rerio*
1	Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	251	33%	9 (3)	36094	7.74	gi|6635240	*Kryptolebias marmoratus*
1	Myosin, light polypeptide 2 , skeletal muscle	243	58%	8 (5)	16639	4.39	gi|47217809	*Danio rerio*
1	Beta-enolase	242	27%	9 (6)	47423	5.99	gi|47210809	*Danio rerio*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
1	Enolase 1 isoform b	206	41%	10 (6)	43939	5.9	gi|226441955	*Gillichthys mirabilis*
1	F-actin-capping protein subunit alpha-1	190	26%	5 (4)	32989	5.41	gi|223647378	*Salmo salar*
1	Fast/white muscle troponin T embryonic isoform	181	16%	6 (2)	33851	5.16	gi|94469903	*Sparus aurata*
1	Sarcomeric mitochondrial creatine kinase-like	174	34%	5 (4)	20475	9.01	gi|292617951	*Danio rerio*
1	Fast/white muscle troponin T embryonic isoform	167	14%	5 (2)	33851	5.16	gi|94469903	*Sparus aurata*
1	Muscle creatine kinase b	163	28%	11 (3)	43100	6.29	gi|157787181	*Danio rerio*
1	Myosin binding protein H-Like	162	13%	6 (2)	54172	6.04	gi|47208602	*Danio rerio*
1	Muscle-type creatine kinase CKM2	161	30%	11 (4)	42694	6.44	gi|21694043	*Oreochromis mossambicus*
1	PDZ and LIM domain protein 7 isoform a	154	33%	8 (2)	24408	6.54	gi|222087993	*Epinephelus coioides*
1	Muscle-type creatine kinase CKM2	153	24%	8 (3)	42980	6.44	gi|21694043	*Oreochromis mossambicus*
1	Muscle-type creatine kinase CKM2	144	26%	11 (5)	42980	6.44	gi|21694043	*Oreochromis mossambicus*
1	Muscle-type creatine kinase CKM2	135	25%	11 (3)	42694	6.44	gi|21694043	*Oreochromis mossambicus*
1	PDZ and LIM domain 7	130	26%	6 (3)	24408	6.54	gi|222087993	*Epinephelus coioides*

Plate	Protein Name	Mascot score	Sequence coverage	Matched peptides (peptides with ion scores)	Mr (kDa)	pI	Accession No.	Species matched in
1	Glyceraldehyde 3-phosphate dehydrogenase	126	21%	6 (4)	36009	8.63	gi|185133678	*Salmo sala*
1	Putative fast skeletal muscle troponin Paralichthys olivaceus	126	60%	7 (2)	10086	9.7	gi|32454282	*Paralichthys olivaceus*
1	Fast/white muscle troponin T larval isoform	117	18%	4 (2)	27389	9.52	gi|94469901	*Sparus aurata*
1	Tropomyoshin1-1	101	18%	8 (3)	33034	4.67	gi|28557116	*Takifugu rubripes*
1	Tropomyosin 3	99	44%	14 (1)	28827	4.76	gi|41393141	*Danio rerio*
1	Myosin-binding protein H-like	95	14%	7 (4)	54172	6.04	gi|47208602	*Danio rerio*
1	Warm temperature-acclimated 65kDa protein	88	69%	4 (3)	7875	7.08	gi|52430344	*Fundulus heteroclitus*
1	Creatine kinase muscle isoform 2	68	17%	5 (3)	42656	6.44	gi|31322099	*Chaenocephalus aceratus*

Note: Proteins highlighted in yellow were identified subsequent to original analysis presented in Abbaraju [Abbaraju, N. (2011) Patterns of protein expression in tissues of the killifish, *Fundulus heteroclitus* and *Fundulus grandis*. Doctoral dissertation, University of New Orleans]

VITA

Mohamed Nazim Boutaghou was born in Algiers, Algeria. In 2000, he left his native country to pursue his college degree in France. He graduated with his Bachelor's degree from the University of Bordeaux I in 2003 with a major in Chemistry and a minor in Biology. He then obtained his Master's degree from the same university in 2005 with a major in Chemistry and a minor in Biology. He pursued as well a degree in business management and graduated from the Business Management Institute of Bordeaux (IAE Bordeaux) in 2007. He moved from France to New Orleans in 2006, where he enrolled at the University of New Orleans to pursue his PhD degree. He joined Dr. Richard B. Cole's research group in January 2007.