A COMPUTATIONAL STUDY ON THE STRUCTURES AND PROTON AFFINITIES OF B_3^+ IONS; PEPTIDE MASS FRAGMENT PRODUCT

A Thesis Submitted to
The Graduate School of Engineering and Science of
İzmir Institute of Technology
In Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

In Molecular Biology and Genetics

By
Seçkin BOZ

July 2015
İZMİR

We approve the thesis of **Seçkin BOZ**.

Examining Committee Members:

_____

Prof. Dr. Nuran Elmacı IRMAK
Department of Chemistry, İzmir Institute of Technology

_____

Prof Dr. Talat YALÇIN
Department of Chemistry, İzmir Institute of Technology

_____

Assoc. Prof. Armağan KINAL
Department of Chemistry, Ege University

31 July 2015

_____

Prof. Dr. Nuran Elmacı IRMAK
Supervisor, Department of Chemistry
İzmir Institute of Technology

_____

Prof. Dr. Ahmet KOÇ
Head of the Department of Molecular Biology
And Genetics

_____

Prof. Dr. Bilge KARAÇALI
Dean of the Graduate School of
Engineering and Science

ACKNOWLEDGEMENT

First of all, I would like to thank my advisor Prof. Dr. Nuran Elmacı for her understanding, patience, and guidance. Without her constant support this thesis would not be written. Also, I am grateful to my committee members; Prof. Dr. Talat Yalçın and Assoc. Prof. Armağan Kınal for their constructive comments, scientific critics and efforts.

Additionally, I would like to express my appreciations to my parents; Semiha Çakal and Bilge Boz. They always believed in me and never stop their constant support.

Following, I would like to show my gratitude to my colleague Sıla Karaca firstly for shorten my orientation period to this lab and secondly his help during this study. Also I want to share my gratefulness for Gün Deniz Akkoç for his programming skills and analytical intelligent.

For financial support, I am pleased to acknowledge the Scientific and Technological Research Council of Turkey (TÜBİTAK) for research project (project no: 111T935). And also thanks to TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA Resources) for the numerical calculations reported in this work.

Lastly, I would like to express my special thanks to my friends Yağmur, Levent, Ozan, Ömer, Mutlu, Talip, Özgür, Özgün, Alev, Yusuf, Emrah and others for their encouragement, support and patience during this graduate work.

ABSTRACT

A COMPUTATIONAL STUDY ON THE STRUCTURES AND PROTON AFFINITIES OF B_3^+ IONS; PEPTIDE MASS FRAGMENT PRODUCT

Mass spectrometry is the tool of choice during most of the proteomics studies to get amino acid sequence. However, unambiguously identifying amino acid sequence from mass spectra is not easy and straight forward task. Deeper understanding is needed to support both existing knowledge and develop newer models on dissociation patterns of protonated peptides and it will help to improve efficiency of current algorithms used in peptide identification.

In this study, the structures of b_3^+ ions and their neutral forms were investigated by using computational methods. First, potential energy surface of b ions are scanned using molecular dynamics simulations and conformer samples are collected. Then, in order to reduce number of conformers, principal coordinate analysis was applied to find and select different structures within the sample. Selected conformers were optimized using density functional theory calculations. Proton affinities of b ions are determined by the energy difference between most stable conformers of the positively charged and neutral peptide fragments.

Different amino acids were used to understand the role of side chain of amino acids on both structures and proton affinities of b_3^+ ions; XA_2^+ where X=N, H, C, Y, D, L and F. The results showed that, b_3^+ ions prefer to have linear oxazolone structure. However, in their neutral states, cyclic structures are relatively far more stable than linear isomers. Histidine display different behavior than other amino acids. Side chain of histidine holds protons and forms stable structures. The energies of cyclic and linear isomers of Histidine containing b ions are close to each other. Histidine containing peptide fragments have larger proton affinity comparing to others. Difference of proton affinities between linear and cyclic conformers varies based on amino acid used. This difference is lower than 10kcal/mol in histidine, asparagine and aspartic acid containing peptide fragments. There is no dramatic position preference of the X-amino acid for the N- or C- terminals or middle position with the exception of Asn and Asp (unlike the center) and Histidine which likes to be at C-terminal .

ÖZET

PEPTİT KÜTLE BÖLÜNME ÜRÜNÜ OLAN B_3^+ İYONLARININ YAPILARI VE PROTON ALMA İSTEKLERİ ÜZERİNE HESAPSAL BİR ÇALIŞMA

Proteinlerin amino asit dizisini elde etmek için kütle spektrometresi sıklıkla kullanılan bir yöntemdir. Fakat kütle spectrumlarından amino asit dizisini net bir şekilde belirlemek zordur. Peptidlerin kimliklerinin saptanması için kullanılan algoritmaların verimini arttırmak adına protonlanmış peptidlerin gas fazındaki kimyasal özelliklerinin araştırılması ve belirlenmesi gerekmektedir.

Bu tez çalışması b_3+ iyonları ve bunların nötr hallerinin hesapsal yöntemlerle incelenmesini amaçlamaktadır. İlk olarak bu iyonların potensiyel enerji yüzeyleri moleküler dinamik simülasyonşarıyla taranmış ve konformer örnekleri toplanmıştır. Daha sonra temel bileşen analizi kullanılarak toplanan örnekler içerisinden yapısal olarak farklılık gösteren izomerler belirlenmiştir. Akabinde bu izomerler Yoğunluk Fonksiyonel Theorisi (DFT) kullanılarak optimize edilmiştir. En kararlı pozitif ve nötr izomerlerin enerjileri farkları kullanılarak proton alma istekleri hesaplanmıştır.

Bu çalışmada farklı amino asitlerin b_3^+ iyonlarının yapılarına ve proton alma isteklerine olan etkisini incelemek için, X=N, H ,C ,Y ,D ,L ve F olmak üzere XA_2^+ peptit parçaları kullanıldı. Sonuçlar b_3^+ iyonlarının oxazolone yapısında olmayı tercih ettiğini göstermiştir. Öte yandan, yüksüz parçacıklarda halkalı yapıların doğrusal yapılara göre çok daha kararlı olduğu tesbit edilmiştir. Histidinin diğer amino asitlerden farklı bir davranış sergilediği ve yan zinciri protonları tutarak kararlı yapılar oluşturduğu gözlenmiştir. Histidin içeren peptit parçaları daha yüksek proton alma isteğine sahip oldukları gibi halkalı ve doğrusal yapıları arasında büyük enerji farkları bulunmamaktadır. Halkalı ve doğrusal yapılar arasındaki enerji farklı kullanılan amino aside göre değişim gösteriyor. Bu fark histidin, aspartik asit ve asparajin içeren peptit parçalarında 10kcal/mol'den azdır. Ayrıca doğrusal yapılar için, kullanılan amino asitlerin çoğunluğunda peptit parçasının C- ya da N- ucunda veya ortada olması önemli bir fark yaratmazken asparajin ve aspartik asit ortada olmamayı histidin ise C- ucuna yakın olmayı tercih etmektedir.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Pages**

LIST OF TABLES

Table **Pages**

CHAPTER 1

INTRODUCTION

Proteins have crucial role in the function of every living organism. For this reason, detailed knowledge about proteins is necessary to understand major metabolic events occurred in the cells. Proteomics is a branch of molecular biology which deals with large set of proteins samples collected from organisms. For proteomics studies, mass-spectrometry is a commonly used tool to determine amino acid sequence of peptides. The details of gas-phase chemistry become important at this point, because in mass-spectrometry peptides are fragmented into smaller molecules in gas-phase. In this chapter, existing understanding about the topic is determined and covered. Detailed literature searches have been done and summarized. Also aim of this research mentioned at the end of this chapter.

1.1. Amino Acids and Proteins

Amino acids are building blocks of proteins. They all contain carboxylic acid, amine group bond to central carbon and variable side chains. There are many types of different amino acids found in living organisms, but 21 of them found in eukaryotes. According to side chain properties amino acids can be divided into groups such as polar, nonpolar, electrically charged. Figure 1.1 is the example of such grouping with the one and three-letter abbreviations of amino acids. In the later part of this thesis these abbreviations can be used inter-changeably.

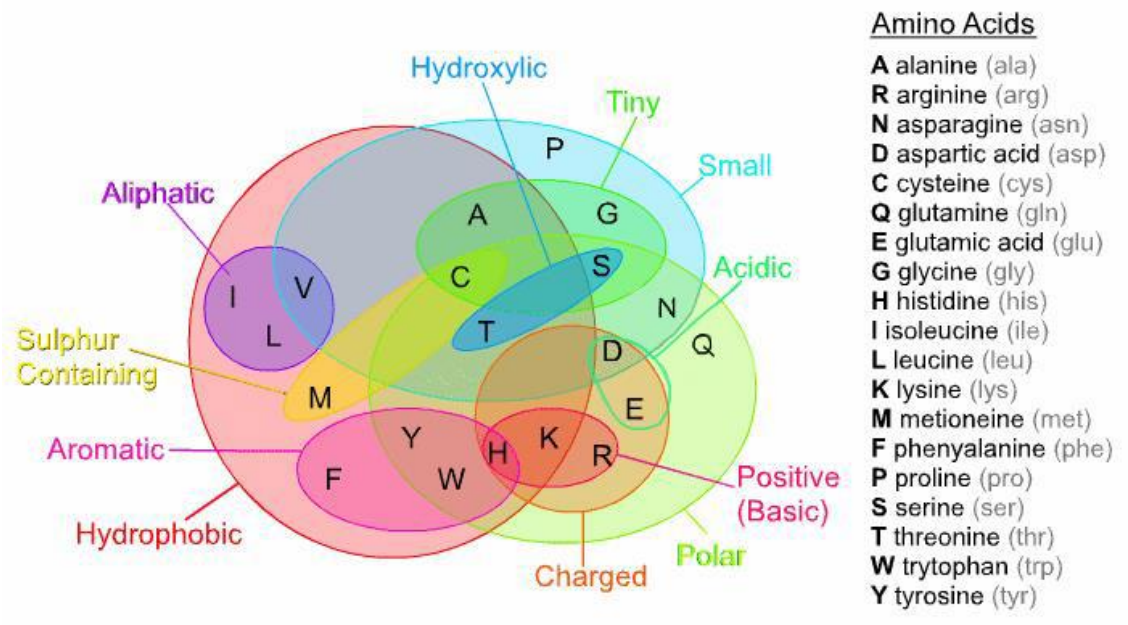

Figure 1.1. Amino acid classification. (Source: http://www.emptycache.com/photographylxft/charged-amino-acids)

Polymerization of amino acids in ribosomes using genetic code creates proteins which are crucial to maintain cellular metabolic events within a cell. Considering all cell in an organism contains same DNA molecule, surprisingly cells in different tissues and organs have different functions. During the flow of genetic information, diversity increases with the help of various mechanisms, such as alternative splicing of mRNAs, post-translational modifications of proteins. For this reason, genomic sequence information alone is not enough to elucidate unknown metabolic systems. At this point protein identification became crucial to get insight about mechanism studied. With the help of recent technological advances, new branch of molecular biology emerges to study total protein content in the cell that called proteomics.

1.2. Mass Spectrometry

During proteomic studies various methods are used to obtain information. One of the most widely used methods involves mass spectrometry. Despite the fact that there are many type of mass spectrometer available in the market and each of them have different working principal we can say that mass spectrometers consist of three main parts; ion source, mass analyzer and ion detector. Figure 2.2 shows basic scheme of generalized mass spectrometer.

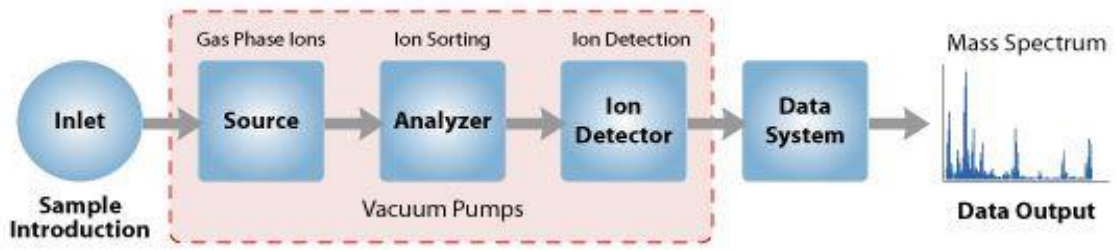

Figure 1.2. Basic scheme of a mass spectrometer. Source produces gas phase ions from liquid peptides. Analyzer is required to separate gas phase ions based on their mass to charge ratio. Ion detector sense resolved ions and records their abundance. (Source: http://www.premierbiosoft.com /tech_notes/mass-spectrometry.html)

Since mass-spectrometry based proteomics is comparably new and rapidly developing technology, it has some drawbacks. One of the biggest mysteries about mass-spectrometry is fragmentation pathways of ionized peptides which involves collision of peptides with noble gases in gas phase. High energy, (de)protonated peptides goes through several chemical reactions and bonds in the backbone and side chains are broken. Nomenclature of these products is based on which part of the peptide proton goes. Figure 1.3 shows this nomenclature.

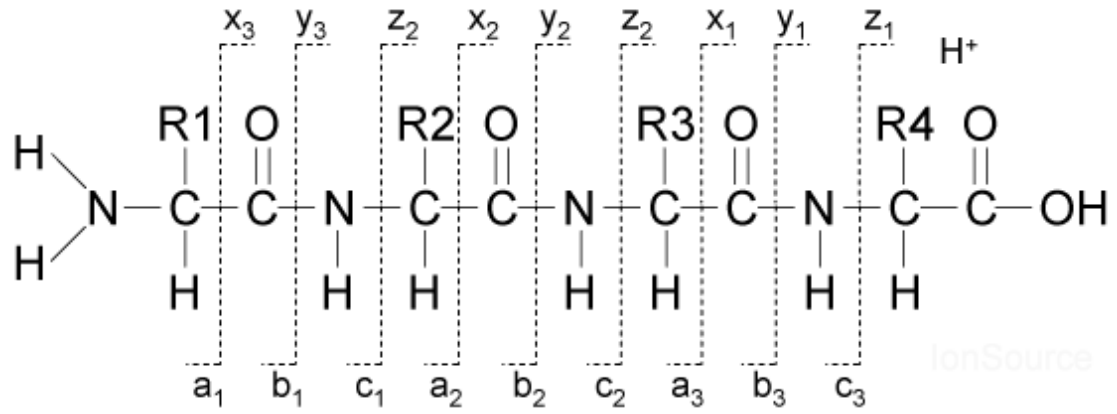

Figure 1.3. Nomenclature of fragmentation of peptides. Letters written above contains C-terminal of peptide and below contains N-terminal. (Source: Harrison, 2009)

In theory, all of the covalent bonds can be broken and peptides can be fragmented to numerous molecules. However, some of the bonds are more likely to be cleaved, and consequent fragments more frequently occur during mass spectrometry studies. For example, peptide bond between carbon and nitrogen is one of these more

frequently broken bonds. For this reason, b and y ion series can be formed often and have vital importance for peptide identification.

1.3. Peptide Identification

In mass spectrometry based proteomics, whole or part of the total protein content in the sample are isolated from lysed cells. These protein contents of cells are digested with enzymes that cleave peptide bond, such as trypsin. Length of these peptides varies between 2 and 20 based on the concentration of enzyme, incubation time and lysine content of the protein. After this, peptide fragment mixture measured by mass spectrometer. Then, recorded mass spectra are used to find out amino acid sequence of the fragment. However, identifying amino acid sequence from mass spectra is not easy task. Different types of approaches exist to solve this problem. Many algorithms commonly used use database search method to match the spectra to specific sequence. There are also algorithms that reveal amino acids sequence directly from mass spectra which called denovo sequencing. Figure 1.3 summarize the flow of procedure during peptide identification.

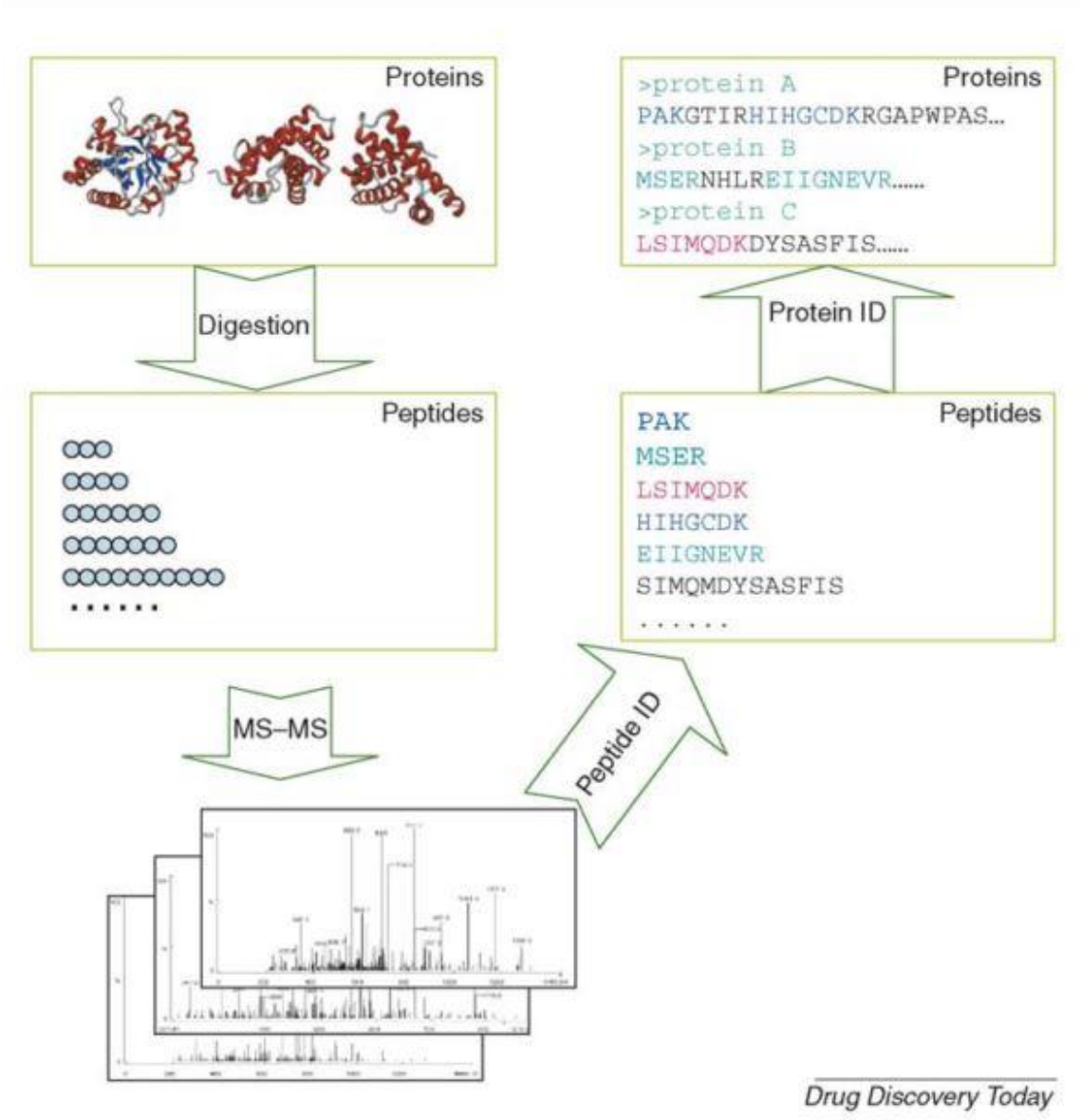

Figure 1.4. Flowchart of mass spectrometry analysis. Identified peptide sequences are later used to determine which protein involves in the sample being studied. (Source: Xu & Ma, 2006)

In this thesis, b ions, which are one of the most abundant ions after fragmentation in mass spectrometry are focused. Behavior of peptides during fragmentation generally depends on amino acid composition, energy and charge. Also, after fragmentation, these ions are still capable of further fragmentations and things going to be unpredictable and make peptide identification from mass spectrum tough procedure for algorithms. For these purpose different models and theories are exist. As an extention of classical models on the fragmentation of protonated peptides, mobile proton model is state-of-art knowledge about the phenomena.

1.4. Mobile Proton Model

MS/MS spectra are produced by using tandem mass spectrometer. Generally with positive ionization mode via electrospray ionization (ESI) technique which is most commonly used ionization technique with matrix-associated laser desorption and ionization (MALDI). In ESI procedure peptides with extra proton are formed. This extra proton usually fastened to N-terminal of the peptide. However, collisions with noble gas increase the energy of the protonated peptide. After this extra proton may move along the whole peptide backbone and held on at amide bonds as well as basic side chains such as arginine and lysine. Figure below illustrate these events.

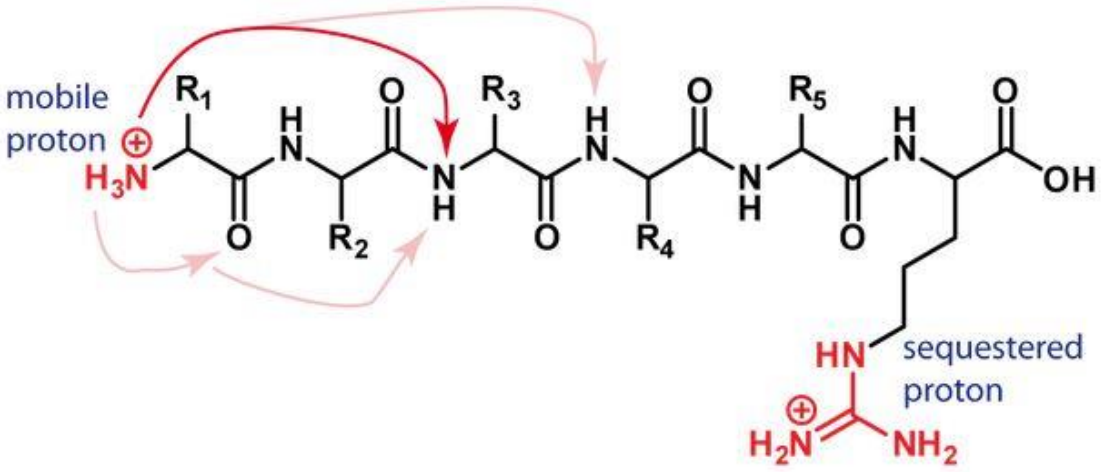

Figure 1.5. Mobile proton model (Source: http://www.lamondlab.com/MSResource/LCMS/MassSpectrometry/mobileProtonModel.php)

This incident is named as mobile proton model. Subsequent fragmentation generates smaller peptide ions based on the probability of extra proton's location. In order to refine algorithms that make peptide identification from MS/MS spectra, we have to deepen our knowledge about gas phase chemistry of protonated peptides.

1.5. Literature Search

Gas-phase chemistry of peptide fragment ions has been explored extensively while past decades. Both experimental and computational studies have been performed in order to get deeper understanding of gas phase chemistry of protonated peptide ions. In one of these studies, fragmentation reactions of protonated peptides explained in

details. In this study, peptide in competition (PIC) model was proposed. According to this model, peptides have several possibilities for further fragmentation and current structure, electronic and chemical properties of peptide are determining factors to which pathways going to happen. Major pathways are b-y pathway, a-y pathways which includes carbon monoxide loss (Paizs & Suhai, 2005). Some pathways includes more information and thus more helpful in order to predict amino sequence from mass spectra. Dissociation of the peptide bonds between central carbon and nitrogen is less likely than other bond in the backbone. For this reason, during mass spectrometry based proteomic studies, probability of the observation of b-y ions are higher comparing to other ions for most of the cases (Bleiholder, Suhai, Harrison, & Paizs, 2011). Also it is mentioned that b ions are capable of involvement to further fragmentation reactions which reduce the intensity of sequence ions produced by b-y pathways. On the other hand, some other ions can be produced and these ions give information that could not be used for prediction of amino acid sequence. These non-sequence ions sometimes observed during experimental studies in special cases. Both experimental and computational works have been done to elucidate these ions in order to get deeper understanding.

Another study specifically focused on b-y pathway and investigates the faith of b ions. According to this study b ions can be fragmented differently according to their size and amino acid composition (A. Harrison, 2009). In this paper, detailed fragmentation reactions of b ions differ in size and composition explained. For b_2^+ ions there are three suggested structures (see Figure below). According to authors, formation of b_2^+ ions can be take place via three ways and each way give results of different structural isomers. First one is cyclic b ion, and called diketopiperazine if b ion contains two amino acid. Second one is oxazolone pathway and produced by when backbone oxygen do nucleophilic attack to central carbon. Last one called imidazole pathway of b ions and rarely occurs. Hence in this thesis we use only cyclic and oxazolone structures of b_3 ions. There are also non-classical b ions whose structure and producing reaction are not similar to these main pathways. Side chain of amino acids and its location effect formation of non-classical b ions.

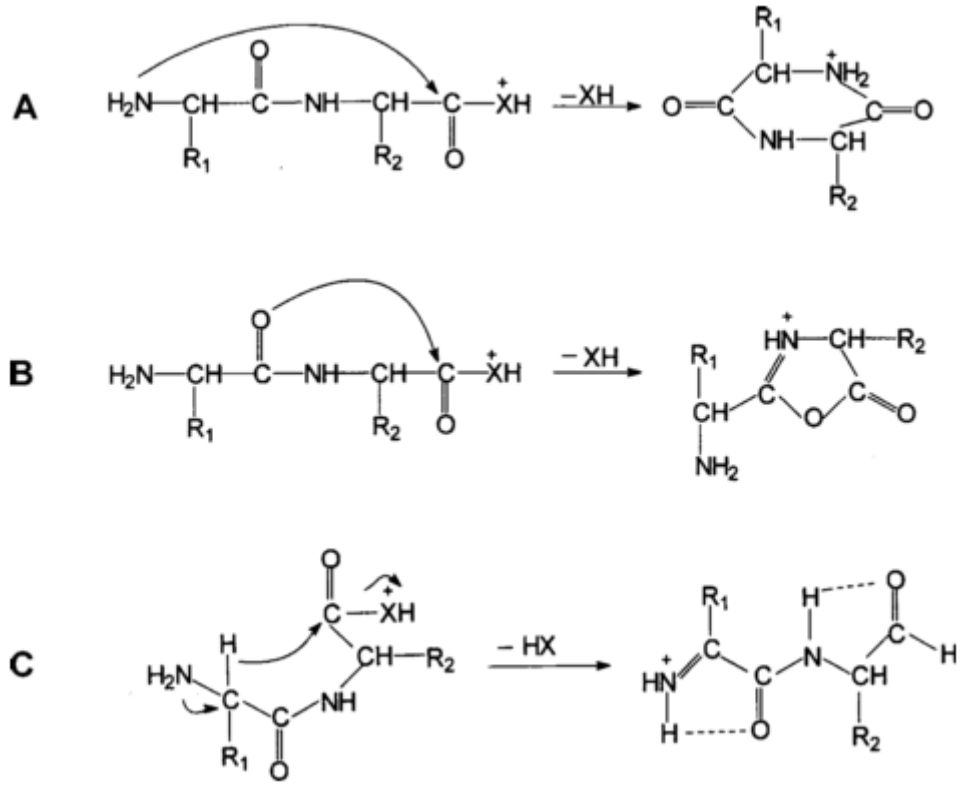

Figure 1.6. Three main pathways that give results on differently structured b_2+ ions. (Soure: A. G. Harrison, 2009b)

There is a competition between pathways A and B. And also final structures of these pathways can be transformed into each other. This reactions called cyclization linearization reactions and results sequence scrambling of protonated peptides (Atik & Yalcin, 2011; Bleiholder et al., 2008). A study in the literature showed that mass spectra of b ions from YAGFL peptide and its derivatives are exactly the same (A. G. Harrison, Young, Bleiholder, Suhai, & Paizs, 2006). Head-to-tail cyclization event of b ions lead to loss of sequence information of mass spectra. Up to now, there is strong evidence that cyclic b_5 ions are exist (Erlekam et al., 2009; Molesworth, Osburn, & Van Stipdonk, 2010). Structure of smaller b ions is not well studied but possibility of cyclization reaction is decreased when peptide ion is smaller due to cis-trans isomerization of peptide bonds.

It has been reported that, if the peptide contains basic residues close to C-terminal, mobile proton take a place around this basic residue and initiate dissociation reaction more frequently comparing to other amino acids. This is called selective cleavage. Amino group at the end of lysine amino acid form cyclic structure with

adjacent central carbon. This cyclic structure known as captolactam ring, and producing pathways is captolactam pathway (Fu, Chen, Xue, Zu, & Fang, 2013).

There are also other exceptions on the structure of b ions. For example, arginine containing peptides prefers selective cleavage near arginine residue which leave Arg side chain close to C-terminal and produce non-classical shape b ions. This five-membered ring called succinic anhydride (Gu, Tsaprailis, Breci, & Wysocki, 2000). Another study systematically investigates fragmentation reactions of tyrosine containing b_3 ions and found location of tyrosine residue significantly affect further fragmentation of these ions (A. G. Harrison, 2009a). Also effect of charge distribution of peptide ions into fragmentation reactions was studied using mass spectrometry and quantum chemical calculations in B3LYP/6-31G(d) level of theory. Oxidation of side chains results in changes in charge distribution and consequently alters pathways for further fragmentation (Sun et al., 2010).

In the literature, large number of research that focused on b_1, b_2 as well as larger b ions can be found. However, there is no computational study on structure of b_3 ions which includes various amino acids. This thesis study is aimed to find most stable structures of b_3+ ions containing alanine, asparagine, aspartic acid, histidine, leucine, tyrosine, phenyl alanine and cysteine. Thus this thesis will show the effect of chemical properties of side chains on the structure of peptide ions. Also location effect of these side chains will be explored. Finding both structures and energies of neutral forms of these ions, yields proton affinity information which has vital importance on the abundance of these ions during mass spectrometry experiments. Addition to this, effect of charge distribution on the structure of peptide ions will be covered. On the other hand, detailed computational work flow to obtain stable conformers of these ions will be described. Consequently, this study unambiguously determines both structure and chemical properties of stable isomers of b_3+ ions and give valuable insights about gas-phase chemistry and further fragmentation reactions.

CHAPTER 2

COMPUTATIONAL METHODS

There are various experimental and computational approach exists in the literature to elucidate properties ionic peptides in gas phase. Experimental methods include mass spectrometry as well as some other methods. Although experimental methods are well studied, straight forward and include no uncertainty about their work flow, computational approaches are not standardized and in every study slightly different methods are used. Computational studies complement results from experimental works and give information about possible reasons of observations. In this chapter you will find information about methods which used in computational chemistry to determine properties of molecules.

2.1. Initial Conformers

All of the linear ion fragments we studied on are in oxazolone formation due to the fact that b ions longer than 2 amino acids tend to form this structure. We used seven different amino acids to see the effect of amino acid type and its position in the sequence on the structure of b3 ions. The location preference of proton on the tripeptide is also analyzed. The proton affinities of those ions are calculated since the one which have higher proton affinity is more likely appear in the mass spectra among the competitive mass fragments. For that purpose different types of amino acids besides alanine (acidic, basic, polar, nonpolar, aromatic, sulfur containing, etc.) were used.

The tri-peptides under the subject of this thesis are XA_2 where X=N, H, C, Y, D, L and F. and A is alanine. Both cyclic and linear conformers were constructed for protonated and neutral tri-peptides. For example initial structures of histidine that His amino acid close to amine group constructed as N-terminal - His-Ala-Ala - C-terminal. Then, histidine group migrated to carboxylic end to see whether there are any significant differences in both proton affinities and the structures due to the position of histidine. Other fragments were constructed same as histidine. Figure below (Figure 2.1) shows

the structures used and their abbreviations which will be used in the later part of this thesis.

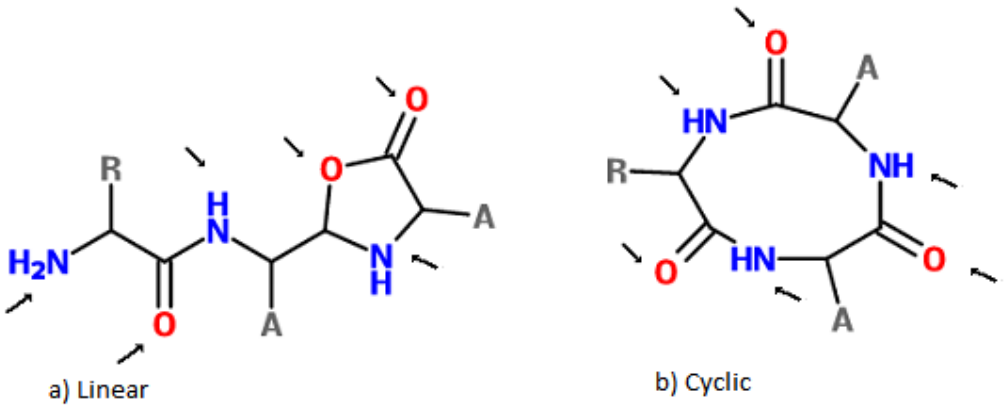

Figure 2.1. Generic representation of initial structures used in this study. In linear structures (a) N-terminal nitrogen, back bone oxygen and nitrogen of oxazolone ring are protonated. They are named as XAA_nterm, XAA_op, and _XAA_oxa respectively. X indicates one letter code of amino acid used. In cyclic conformers (b), only oxygens are protonated. Starting from R group and turning clock-wise around backbone cycle, we named protonated isomers as XAA_cyc_O1, XAA_cyc_O2, XAA_cyc_O3.

After constructing initial structure, possible protonation sites of the molecule determined. Oxygen and nitrogen atoms have this potential. In order to determine which protonation sites are relatively stable than other sites, all possible protonation sites are considered. These sites are oxygen and nitrogen atoms in the backbone of peptide and side chains if there are any available protonation site exists. Consequent findings showed that backbone nitrogen and oxygens in oxazolone rings in linear conformation have relatively much more unstable comparing to other protonation sites and conformer with these protonation sites are not studied further. Figure above also shows generic representation of typical peptide investigated in this work. Protonation sites are indicated by arrows.

To avoid any problems caused by initial constructs and to find their charge distribution each conformer was subject to optimization. During this optimization DFT method was used which will be mentioned later part of this chapter. Optimized initial conformers were used to generate conformer samples from potential energy surface by using molecular dynamics simulation.

2.2. Molecular Mechanics

Molecular mechanics is a technique that is used to optimize the structure of molecules by taking into account of potential energy function. Force field is main composition of this method and contains structural parameters and equations to calculate forces acting on particles. The technique based on classical rules of physics. Conformational energy of a molecule is changing relative to some factors. These factors are shown in the following equation.

$$E = E_{(\Theta)} + E_{(r)} + E_{(\Phi)} + E_{(vdW)} + E_{(el)} \quad (2.1)$$

Potential energy of a conformer is divided into two main categories; covalent and non-covalent interactions. Covalent interactions involves bond stretching ($E_{(r)}$), angular bending ($E_{(\Theta)}$) and dihedral angle torsion ($E_{(\Phi)}$). Non-covalent interactions can be expressed as van deer Waals ($E_{(vdW)}$) and electrostatic terms (Rappé & Casewit, 1997).

During molecular dynamics simulation which will be explained later part of this thesis the forces such as Lennard-Jones potential which tells two atoms attract or repel each other depending on the distance between them or electro static forces that occurs if the atoms are positively or negatively charged must be known and defined. For this purpose, each molecular dynamic simulation needs a proper force definition suitable for simulated molecules and atoms. Type of force field effect the quality of simulation and need to be compatible (Engler, Andose, & Schleyer, 1973).

2.3. Molecular Dynamics Simulation

A recent research article evaluate performance of conformational search approaches like classical molecular dynamics, simulated annealing, Monte-Carlo (MC), energy leveling, and basin hopping (Grebner, Becker, Stepanenko, & Engels, 2011). In this thesis study, we try MC, classical molecular dynamics, and simulated annealing methods to search conformational space. Consequent evaluation of these methods show us molecular dynamic with simulated annealing have better performance considering computational time, ability to find conformers with lower energy.

Molecular dynamics simulation is a method to simulate and visualize behavior of molecules in computer environment. It uses Newton's rules of movement. Following equation indicate Newton's second law:

$$F_i = m_i a_i \tag{2.2}$$

F_i is the force that acts upon a particle that possesses m_i mass and a_i is acceleration of that particle. Taking into account that acceleration is changes in velocity over time and the velocity is total displacement in given amount of time and considering acceleration is not changing during each time step; the equation can be re-written as below:

$$F = m \frac{dv}{dt} = m \frac{d^2x}{dt^2} \tag{2.3}$$

As equation 2.3 describe, the method is time dependent and locations of particles are calculated using their first locations and velocities in each step. In order to find final velocity consequent equations is used:

$$v_f = a.t + v_i \qquad x_f = v.t + x_i \tag{2.4}$$

Combining these two formulas give us:

$$x_f = a.t^2 + v_i.t + x_i \tag{2.5}$$

Further, using potential energy formula, the acceleration can be described as the equation below with regard to position of particle, x:

$$a = - \frac{1}{m} \frac{dE}{dx} \tag{2.6}$$

All of these equations show that, initial velocities, initial positions and acceleration are needed to calculate trajectories of given N-body systems which in this case molecules. Since average velocity of a molecule have a relationship to temperature,

Gaussian or Maxwell-Boltzmann distribution could be used to generate initial velocities according to temperature (D C Rapaport, 1997). This relationship can be expressed as formula below: where v_x^2 is the average translational kinetic energy, k_b is Boltzmann constant and T is the temperature (Li, 2005).

$$\frac{1}{2} m v_x^2 = \frac{3}{2} k_b T \tag{2.7}$$

Molecular dynamic simulation is performed via numerically solving Newton's equations of motions which was expressed above. Figure below shows simplified algorithm for molecular dynamics simulation.

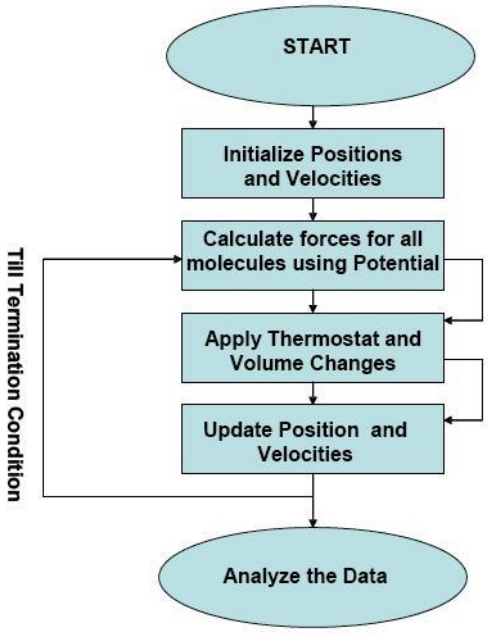

Figure 2.2. Simplified algorithm for molecular dynamic simulation. Initial positions and velocities are main input of the algorithm. Forces are calculated by force field used. Temperature and pressure regulation is made with user defined parameters. Final position and velocities are determined and used as in first step. (Source: http://phycomp.technion.ac.il/~pavelba/Comp_Phys/Project/Project.html)

The molecule to be simulated is placed in box-like boundaries. These box-like boundaries are located near to each other, thus molecule flow freely through space otherwise would hit the edge of these boundaries (D. C. Rapaport, 1999). Figure 2.2 give an idea about the concept of periodic boundaries.

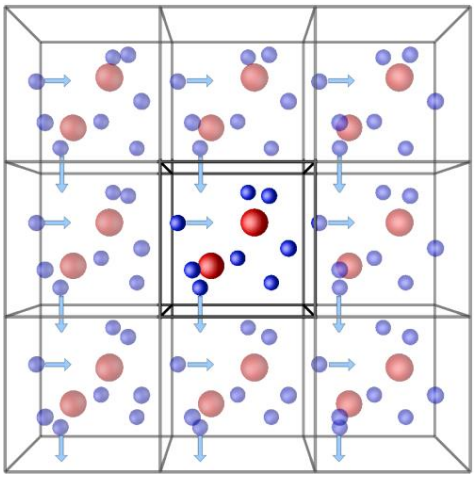

Figure 2.3. Diagrammatic figuration of periodic boundary conditions. There are exact copies of molecules in each box. (Source: http://people.cst.cmich .edu/petko1vg/isaacs/phys/pbc.html)

Amber11 (Case et al., 2011) program is used to scan through potential energy surface and generate low-energy conformers. Potential energy surface or landscape used to describe energy of the system via corresponding configuration, in this case molecules (see Figure 2.3). In this program, molecular dynamic simulation of b_3+ ions was carried out. Force field that specialized for protein backbone dihedral potential was used during simulation. Force field named as ff99sb and found in Amber11 program. However, this force field is not sufficient for protonated oxygen and nitrogen species. Thus, we modified the force field for these cases.

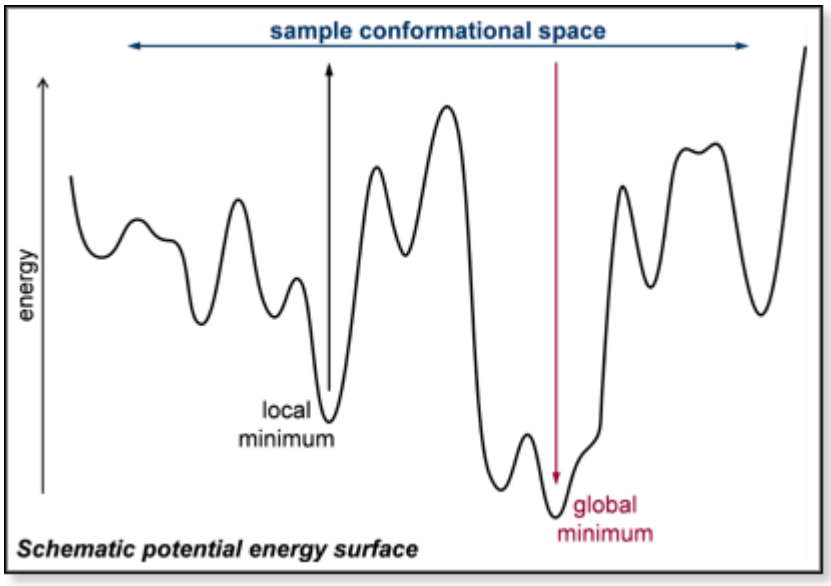

Figure 2.4. Two dimensional representation of potential energy surface. Potential energy is higher in upper spaces. Hills in the figure show conformations that have relatively high energy like transition states. (Source: https://labs.chem.ucsb.edu/bowers/michael/theory_analysis/mm/sampling.shtml

Simulation starts from room temperature and system is gradually heated to increase total energy of the system. Heating part is 12 pico-seconds long. Temperature rises to 1500 K during heating stage. High temperature values allow molecules to move freely without be fastened by energy barriers on potential energy surfaces. During 2 nano-seconds long simulation; samples are collected in every 5 ps intervals. Coordinates of the each atom of the molecule in three dimensional space are saved and used in cooling procedure. This method is known as the simulated annealing and is used to move molecules on potential energy surface to find local minimums. Cooling stage is 16 ps. However, in the first 4 ps of this cooling procedure temperature rapidly decrease to 0 K. This allows molecules to settle lowest energy states in their corresponding location of potential energy surface. After this, frozen molecules are again heated to 100 K and allowed to move while 5 ps. Lastly, during 4ps long step, total energy of the system lowered to 0K. This is the part of the simulated annealing procedure and called as echo cooling. This is applied to find local minimum of each conformer on their potential energy surfaces. Each stage of the molecular dynamics simulation was carried out with 1 fs time step.

Consequently, 4000 conformers are obtained for each isomer. Due to this number is much higher to optimize all of them using quantum chemistry calculations in

practice, statistical sampling method named principal coordinate analysis was used to reduce number of samples.

2.4. Principal Coordinate Analysis

Several methods are used to reduce sample size for further optimization in the literature. One of these methods is clustering conformers based on their dihedral angles in peptide backbone (Paizs, Suhai, Hargittai, Hruby, & Somogyi, 2002). Another method is to use principal component analysis. There are a few examples in the literature that use this approach (Becker, 1998). In this thesis, principal coordinate analysis method was used to distinguish significantly different conformers among whole set of conformers comes from molecular dynamics simulations.

Principal coordinate analysis (PCoA) is a statistical method to visualize similarities of each individual (objects) in the population. The method is also known as classical multi-dimensional scaling (MDS). Idea behind PCoA is to represent system as dissimilarity matrix and reducing the number dimensions in this matrix. Use of PCoA is practical when dataset have many objects and attributes that describe characteristics of each object (Jolliffe, 2002). The method transforms large data sets which contain correlated variables among objects into fewer uncorrelated variables. These uncorrelated variables also carry information about the system and represent whole set of original variables using lower dimensions. Historically, the method is first proposed by Pearson in 1901. Since then application of this method finds many areas today's science and information technologies, such as biology, medicine, chemistry, geology, behavioral and social sciences, marketing and banking information system (Dunteman, 1989).

After collecting samples from molecular dynamic simulation, we propose a novel implementation of usage of PCoA to reduce sample size in numbers. 4000 conformers were collected from molecular dynamic simulation with simulated annealing and this size is computationally extremely expensive in practice to optimize all of them using density functional theory calculations. For this reason, PCoA is used to visualize multi-dimensional information comes from different conformations of b_3 ions.

Procedure is begun with calculating dissimilarity matrix. To calculate dissimilarity matrix, variables that defines individuals are used. These variables are

coordinates of each atom in three dimensional space within the molecule examined in our case. PCoA method utilize *nxn* matrix, where *n* is the total number of objects which is 4000 in this study. With the help of this matrix PCoA try to find out objects whose variables are similar to each other. Calculating dissimilarity matrix was done by using the formula below:

$$D_{ij} = \sum_{r=1}^{p} (X_{ir} - X_{jr}) \tag{2.8}$$

Where *i* and *j* are conformation indexes, *r* is the index of variables, *X* is the matrix that contains variables and *D* represents the value of dissimilarity between i^{th} and j^{th} conformers. Four different *X* matrices were constructed from original data. These are Cartesian coordinates of each atom, relative atomic distances, bond angles and dihedral angles. Among them relative atomic distances indicates more alteration on conformers, and was used as similarity measurement. In this measurement, distance between each two atoms is recorded as a variable. So, we have *p* variables in the matrix *X*. *N* is the total number of atoms (Elmaci & Berry, 1999).

$$p = \frac{N(N-1)}{2} \tag{2.9}$$

After that procedure is continued as follows:

- Using the element of dissimilarity matrix, d_{ij}, construction of matrix **A** whose elements are $-d_{ij}/2$;
- Centralization of matrix **A** using following equation and defining matrix **B** with acquired elements:

$$b_{ij} = a_{ij} + a_{i.} + a_{j.} + a_{..} \tag{2.10}$$

where b_{ij} and a_{ij} are corresponding elements of matrix **B** and **A**, $a_{i.}$ and $a_{j.}$ are average of values of i^{th} row and j^{th} column in matrix **A**.

- Diagonalization of matrix **B** and finding its eigenvalues and eigenvectors take place. Corresponding eigenvectors of each eigenvalues are normalized.

The biggest eigenvalue represents the highest alteration on original data set and contains most of the information about the system. One can use as much as eigenvalue they need; however two or three is enough to present data graphically. Figure below is distribution of two eigenvectors which correspond to first two eigenvalues. Each dots are belong to one conformer. Closer the dots are similar the structure is. Color and area of the dots represents potential energy of these conformers comes from molecular dynamic simulation with simulated annealing method.

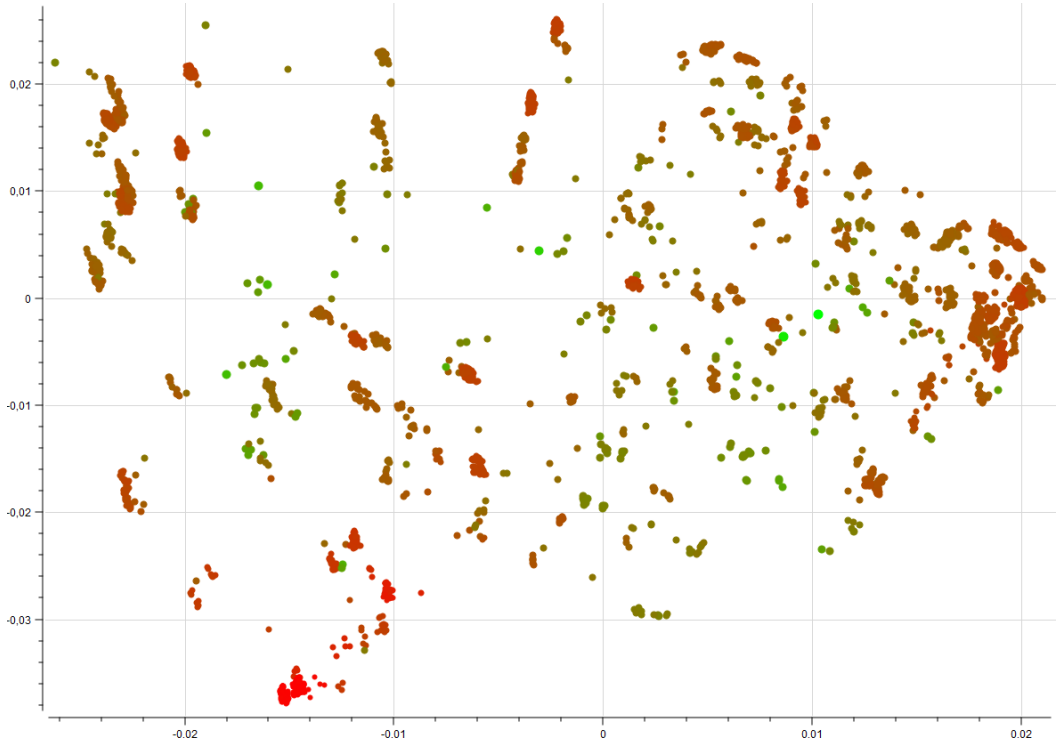

Figure 2.5. Distribution of first two eigenvalues among two dimensional space. Colors and diameter of dots indicate potential energy of the conformers. Red and smaller dots have lower energy comparing to green and big dots.

After obtaining first two dimensions and plotting the graph above, we simply divide this graph into rectangles as shown. Collecting minimum energy conformers from each rectangle allowed us to reduce sample size without losing important conformers for further refinement using density functional theory

2.5. Density Functional Theory

Density functional theory (DFT) is used to calculate electronic structure of atoms and molecules. It is based on the solution of many-electron time-independent Schrödinger equation:

$$\hat{H}\,\Psi = [\ \hat{T} + \hat{V} + \hat{U}\ ]\,\Psi \qquad (2.11)$$

Where $\hat{H}$ is Hamiltonian operator which corresponds to total energy of the system, $\hat{T}$ is the kinetic energy, $\hat{V}$ is the potential energy that comes from charge of nuclei and $\hat{U}$ is repulsion energy that created by electron-electron interaction (Foresman & Frisch, 1996).

The method was first proposed at 1964 by Hohenberg and Kohn. Their theorem has idea of external potential of molecules can be determined by electron density. One year later Kohn and Shan make first implementation of DFT and present it to use of scientific community. Functional which means function of a function using to calculate electron density in quantum mechanics proposed by Kohn and Shan is similar to Hatree-Fock (HF) method. However, DFT is faster than HF calculations because it uses three dimensional function rather than quadratic function used in HF method (Young, 2004). Since then lots of refinement have been making on its original form, DFT has proved its usefulness in computational sciences such as physics and computational chemistry.

Density functionals can be divided into various categories. Xα is the most basic one among them and it only considers electron exchange not correlation. There are other calculations as well, such as hybrid method B3LYP which stands for Becke, 3 parameter, Lee, Yang, Parr exchange. Hybrid methods generally give more accurate results comparing to gradient-corrected method since they include functionals to calculate exchange integrals (Orio, Pantazis, & Neese, 1983).

In this thesis study, Gaussian09 program was used for final geometry optimization with DFT (Frisch et al., 2009). In most of the examples in literature for the structure determination of b ions, density functional theory found suitable (Chass, Marai, Setiadi, Csizmadia, & Harrison, 2004; Chen & Turecek, 2005; Tang, Fang, Harrison, & Csizmadia, 2004). Hence, DFT was used to optimize selected structures and obtain their final energies.

CHAPTER 3

RESULTS AND DISCUSSIONS

In this chapter, optimized final structures of b_3^+ ions and their neutral forms was evaluated. All of the final optimizations have been done at B3LYP/6-31G+(d,p) level of theory. Also zero-point energies and charge distributions was calculated. Seven amino acids was investigated in this study; asparagine, aspartic acid, leucine, histidine, cysteine, tyrosine, and phenylalanine. Structure of cyclic and linear isomers of both b_3^+ ions and neutral species has been determined. Also charge distribution of these peptide mass fragments has been calculated.

3.1. b_3^+ Ions

Asparagine which is a polar amino acid has carboxamide (-$CONH_2$) side chain group. b_3^+ ions of NAA, ANA, and NAA was protonated from possible sites which are N-terminal nitrogen, nitrogen atom of oxazolone ring, and backbone oxygens and nitrogens. Also side chain of asparagine can be protonated. Table 3.1 shows electronic and zero-point corrected energies (ZPE) of different configurations with protonation sites as both atomic unit (au) and relative energies (kJ/mol). ID of these conformers indicates sequence number from molecular dynamic simulation.

Table 3.1. Electronic Energies and zero-point corrected relative energies of NA_2

P. Site	ID	Energy(au)	ZPE(au)	Rel E. kJ/mol	Rel ZPE kJ/mol
NAA_oxa	3745	-911,1014914	-910,812209	0,5	0,0
AAN_nterm	2808	-911,1016721	-910,810472	0,0	4,6
NAA_op	3427	-911,0959836	-910,808053	14,9	10,9
Cyc_asn_op	2307	-911,0958044	-910,807062	15,4	13,5
NAA_nterm	1880	-911,0966121	-910,806717	13,3	14,4
ANA_oxa	1811	-911,0937338	-910,805371	20,8	18,0
AAN_op	2376	-911,0940851	-910,805302	19,9	18,1
ANA_op	2367	-911,0922015	-910,804568	24,9	20,1
AAN_oxa	1881	-911,0911909	-910,802734	27,5	24,9
ANA_nterm	2555	-911,0928185	-910,802323	23,2	26,0
NAA_asn_op	3129	-911,0867105	-910,797806	39,3	37,8
AAN_asn_op	62	-911,0845653	-910,795697	44,9	43,4
Cyc_o3	3787	-911,0826547	-910,793689	49,9	48,6
Cyc_o1	2847	-911,0816747	-910,791367	52,5	54,7
ANA_asn_op	3107	-911,0744003	-910,786961	71,6	66,3
Cyc_o2	189	-911,0770474	-910,786353	64,7	67,9

Figure below shows the structure of the most stable three linear and one cyclic isomers of NA_2. Tubular representation is used in order to avoid graphical complexity. Proton is shown as white sphere. Closest atom to proton determined as protonation site. Relative energies are represented as kJ/mol. Polar contacts, such as hydrogen bonding, are displayed as yellow dots.

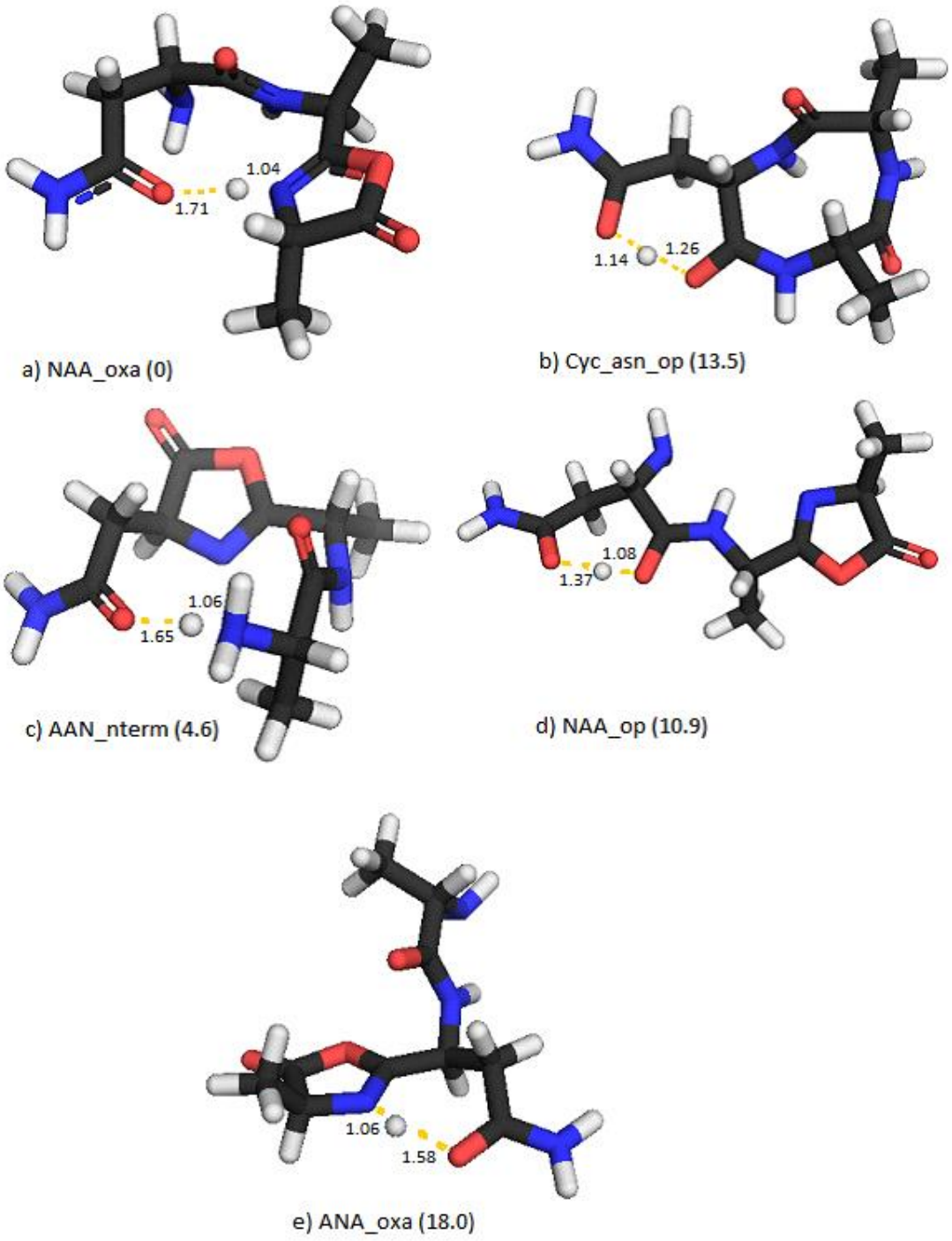

Figure 3.1. Structures of b_3^+ ions of NA_2

The most stable isomer of b_3+ ion of NA_2 is nitrogen of oxazolone ring protonated one. In this configuration, oxygen atom of aspartic acid residue closes to N-terminal of peptide ion make hydrogen bond with protonated nitrogen of oxazolone ring. Thus, circular like structure is formed (Fig 3.1.a). In second stable isomer N-

terminal protonated nitrogen and oxygen of asparagine side chain close to carboxylic end form hydrogen bonds (Fig 3.1.c). 4.6 kJ/mol energy differences exist between NAA_oxa and AAN_nterm isomers. The zero point energy correction changed the order of stability of these two conformers. NAA_op which refers to backbone oxygen protonated conformer is the third most stable one and have 10.9 kJ/mol energy differences from most stable isomer. Oxygen of asparagine side chain protonated cyclic isomer has only 13.5 kJ/mol higher energy. However, linear isomers of this protonation site have much higher energies; 37.6, 43.4, 66.3 kJ/mol. Also protonation site of these two conformers almost have the same structure (Fig 3.1.b, d) and this structure creates 7-membered ring with proton. Two oxygen atoms get close to each other and proton located between them including other four carbon atoms. This structure is also observed in other peptide fragment ions.

Aspartic acid has carboxylic (-COOH) acid in its side chain. Protonated DAA, ADA, and AAD isomers and cyclic forms have been investigated. Side chain of aspartic acid has low pK_a (3.9) and can be found as negative ion depending to chemical environment. Table below shows energies with zero point(ZP) correction of DA_2 isomers.

Table 3.2. Electronic Energies and zero-point corrected relative energies of DA_2

P. Site	ID	Energy(au)	ZPE(au)	Rel E. kJ/mol	Rel ZPEkJ/mol
DAA_oxa	2729	-930,9560382	-930,679694	2,7	0,0
AAD_nterm	2158	-930,9570707	-930,678282	0,0	3,7
AAD_op	3003	-930,9520339	-930,677002	13,2	7,1
ADA_oxa	194	-930,951749	-930,676184	14,0	9,2
DAA_nterm	2765	-930,9536058	-930,675701	9,1	10,5
ADA_op	3270	-930,9503421	-930,675296	17,7	11,5
AAD_oxa	2817	-930,949871	-930,674258	18,9	14,3
ADA_nterm	2494	-930,9524085	-930,673618	12,2	16,0
Cyc_o3	3937	-930,9482554	-930,670138	23,1	25,1
Cyc_o1	567	-930,9437341	-930,665597	35,0	37,0
Cyc_o2	833	-930,9362206	-930,657805	54,7	57,5

Figure below shows the structure of most stable three linear and one cyclic isomer of DA_2.

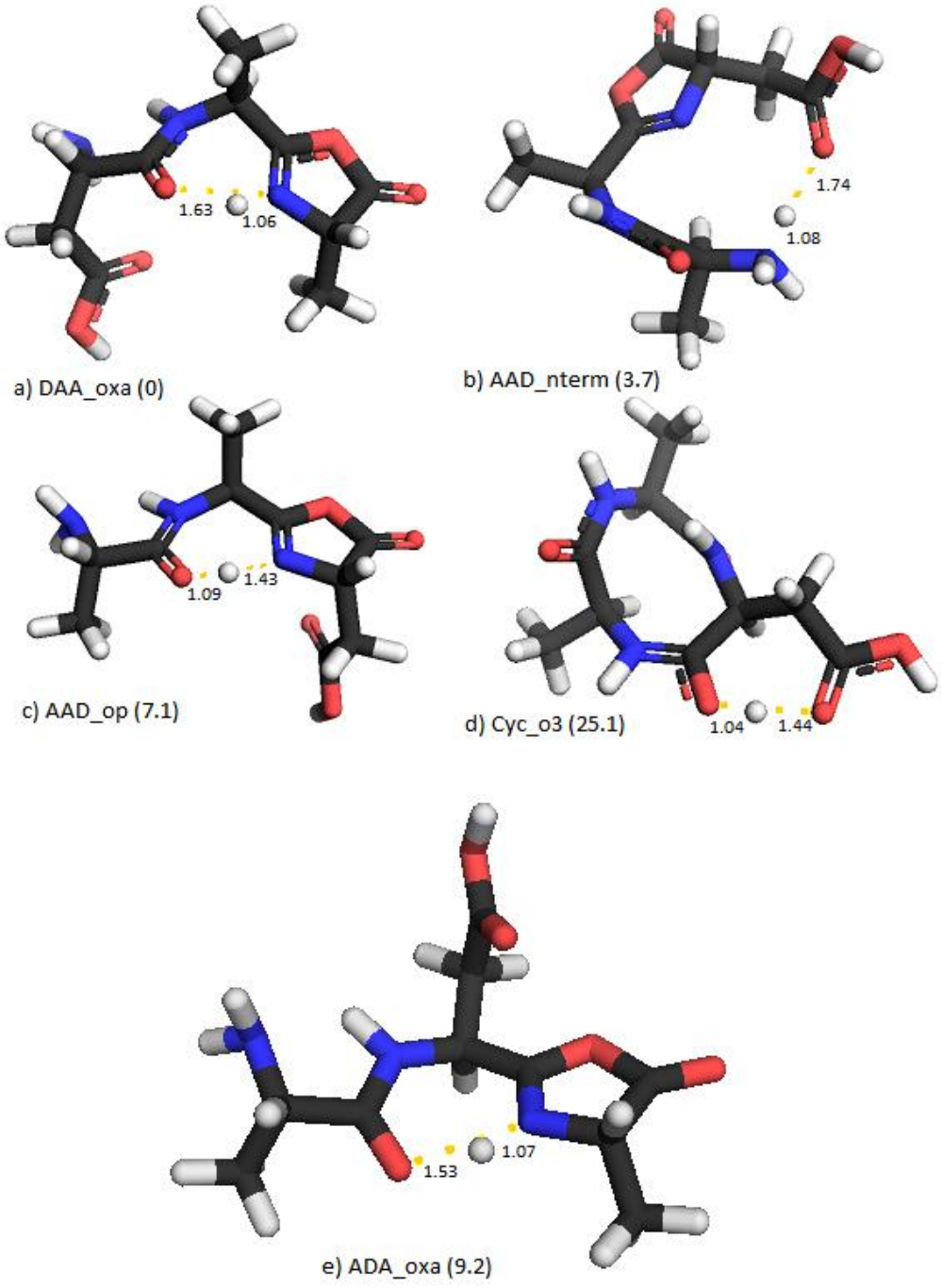

Figure 3.2. Structures of b_3^+ ions of DA_2

The lowest energy isomer of DA_2 is DAA_oxa and has only 3.7 kJ/mol energy difference from AAD_nterm. Only N-terminal protonated ion displays circular like structure. DAA_oxa and AAD_op have similar appearance and proton located in the middle of conformers. Backbone oxygen protonated ion of linear conformer with side chain of aspartic acid close to carboxylic end is third stable isomer and its relative energy is 7.1 kJ/mol. The first three isomers have similar trend comparing to

asparagine. However, side chain of aspartic acid has no possible protonation site and cyclic isomers have much higher relative energies; 25.1, 37.0, 57.5 kJ/mol. Same as asparagine, 7-membered ring with proton occurs in cyclic isomer.

Leucine have branched aliphatic side chain and considered as hydrophobic. As linear conformers LAA, ALA and AAL combinations have been constructed and possible protonation sides were protonated. Leucine has no possible protonation sites in its side chains. Table 3.3 shows zero-point corrected energies of leucine containing b_3^+ ions.

Table 3.3. Electronic Energies and zero-point corrected relative energies of LA_2

P. Site	ID	Energy(au)	ZPE(au)	Rel E. kJ/mol	Rel ZPE kJ/mol
AAL_oxa	503	-860,3272128	-859,981795	0,0	0,0
ALA_oxa	3183	-860,3265959	-859,981468	1,6	0,9
AAL_op	438	-860,3244604	-859,981321	7,2	1,2
ALA_op	3303	-860,3248344	-859,980705	6,2	2,9
LAA_oxa	1765	-860,3244968	-859,979702	7,1	5,5
LAA_nterm	914	-860,3233439	-859,976828	10,2	13,0
LAA_op	2938	-860,321094	-859,975215	16,1	17,3
AAL_nterm	3383	-860,3205277	-859,974097	17,6	20,2
ALA_nterm	1861	-860,3192663	-859,972588	20,9	24,2
Cyc_o1	411	-860,3086799	-859,960795	48,7	55,1
Cyc_o3	215	-860,3062217	-859,957367	55,1	64,1
Cyc_o2	164	-860,3050052	-859,957073	58,3	64,9

Figure below shows the structure of most stable three linear and one cyclic isomer of LA_2.

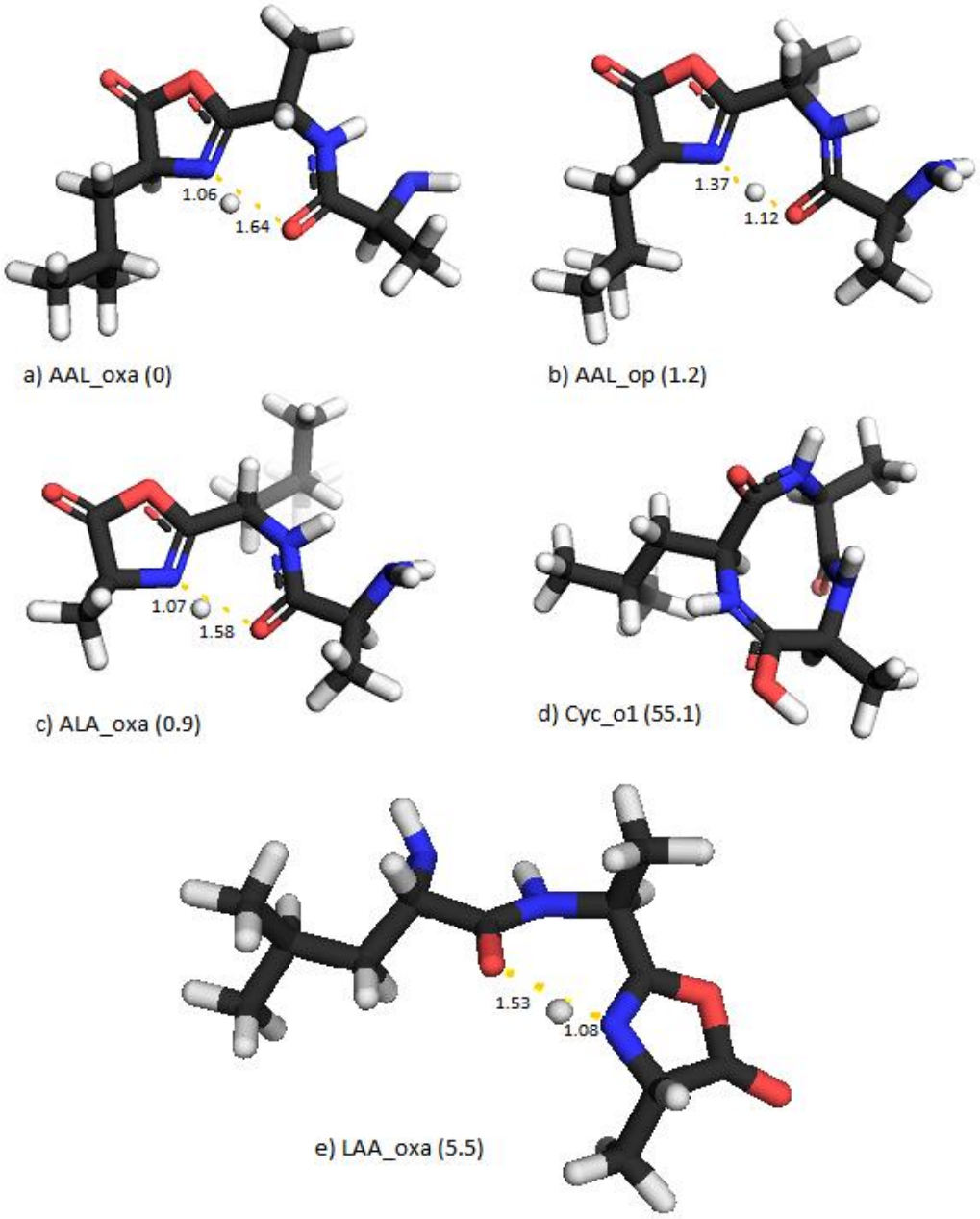

Figure 3.3. Structures of b_3^+ ions of LA_2

Nitrogen of oxazolone ring protonated AAL_oxa conformer is most stable isomer among others. Also side chain of leucine prefers stay close to carboxyl terminal of peptide fragment. First four lowest energy isomers are energetically similar; only 2.9 kJ/mol energy differences exist between them. Configurations of atoms which are around protonation sites of these four conformers are surprisingly similar. Proton is located between nitrogen and oxygen and produce stable structure like 7-membered ring. On the other hand, 55.1 kJ/mol energy differences are much higher between lowest energy linear and cyclic isomers comparing to other amino acids. Possible reason of this

difference can be not containing polar side chain of leucine which balance the charge derived from proton. Because both asparagine and aspartic acid have such side chains, and have relatively lower energy barrier between cyclic and linear isomers. N-terminal protonated isomers are higher energy (13.0 kJ/mol) which mean this position and protonation site are not preferred in leucine containing peptide ions.

Histidine side chain has imidazole ring which shows aromatic and basic chemical properties. Histidine side chain can be protonated from nitrogen of imidazole ring. Only one of two nitrogen atoms in the ring can be protonated, because one of them has one proton as default. Energies of various HA_2 conformers are displayed in the Table 3.4.

Table 3.4. Electronic Energies and zero-point corrected relative energies of HA_2

P. Site	ID	Energy(au)	ZPE(au)	Rel E. kJ/mol	Rel ZPE kJ/mol
AAH_his	3949	-967,4417539	-967,127212	0,0	0,0
Cyc_his	3986	-967,4414763	-967,12489	0,7	6,1
AHA_his	1151	-967,4366071	-967,122958	13,5	11,2
HAA_his	2673	-967,4365045	-967,122855	13,8	11,4
AAH_nterm	3730	-967,425831	-967,110809	41,8	43,1
AHA_nterm	1709	-967,4120067	-967,096645	78,1	80,3
AAH_oxa	1376	-967,4063149	-967,093758	93,0	87,8
AAH_op	1197	-967,4058697	-967,093483	94,2	88,6
HAA_nterm	3162	-967,4075395	-967,092488	89,8	91,2
HAA_oxa	3831	-967,4061663	-967,092402	93,4	91,4
AHA_oxa	1014	-967,4043054	-967,091489	98,3	93,8
AHA_op	3806	-967,403384	-967,09127	100,7	94,4
Cyc_o1	3961	-967,406424	-967,09431	92,8	86,4
HAA_op	770	-967,4022168	-967,090825	103,8	95,5
Cyc_o2	3963	-967,4023442	-967,086629	103,5	106,6
Cyc_o3	2358	-967,3988974	-967,083358	112,5	115,1

Figure 3.4 shows the structure of most stable three linear and one cyclic isomer of HA_2.

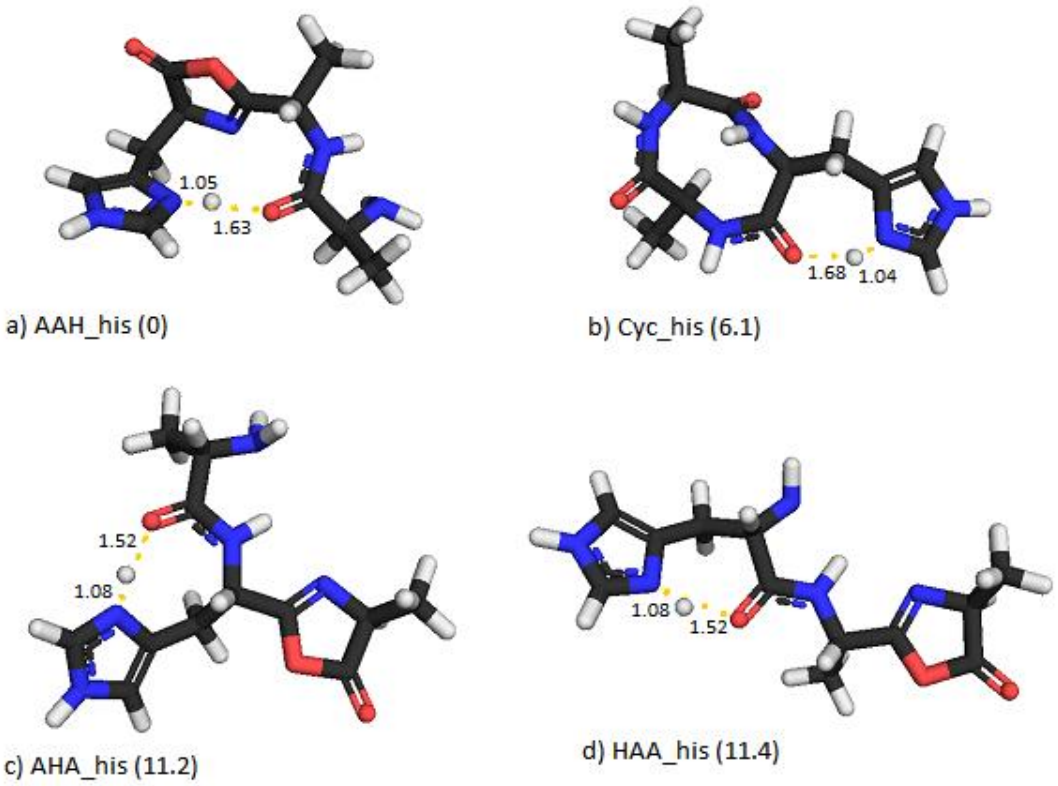

Figure 3.4. Structures of b_3^+ ions of HA_2

Histidine protonated isomers are far more stable than other protonation sites. It may be because of higher proton affinity of histidine side chain. AAH_his conformer has lowest energy and has 6.1 kJ/mol differences between Cyc_his conformer. In all of histidine protonated conformers, nitrogen of protonated histidine makes hydrogen bond with backbone oxygen. Also there are polar contacts with nitrogen of oxazolone ring in AAH_his. After histidine protonated isomers, N-terminally protonated AAH_nterm and AHA_nterm isomers take place with 43.1 and 80.3 kJ/mol relative energies accordingly.

Tyrosine contains aromatic ring in their side chains. Tyrosine have –OH group at the ring. This group play important role in biological functions of tyrosine. Tyrosine has no protonation site in its side chains. b_3^+ ion combinations of tyrosine have been constructed. Table 3.5 indicates electronic energies including zero point corrections as well as relative energies of YA_2.

Table 3.5. Electronic Energies and zero-point corrected relative energies of YA_2

P. Site	ID	Energy(au)	ZPE(au)	Rel E. kJ/mol	Rel ZPE kJ/mol
AAY_oxa	3450	-1048,667199	-1048,32123	0,0	0,0
AYA_oxa	2443	-1048,665922	-1048,32038	3,4	2,2
AAY_op	1848	-1048,664118	-1048,31983	8,1	3,7
YAA_nterm	1943	-1048,666348	-1048,31871	2,2	6,6
YAA_oxa	1218	-1048,663858	-1048,31812	8,8	8,2
YAA_op	2073	-1048,663201	-1048,31788	10,5	8,8
AAY_nterm	2698	-1048,665277	-1048,31666	5,0	12,0
AYA_nterm	3346	-1048,660739	-1048,31282	17,0	22,1
Cyc_o3	735	-1048,651872	-1048,30312	40,2	47,6
Cyc_o2	2945	-1048,645103	-1048,29664	58,0	64,6
Cyc_o1	2356	-1048,64521	-1048,29654	57,7	64,8

Figure below shows the structure of most stable three linear and one cyclic isomer of YA_2.

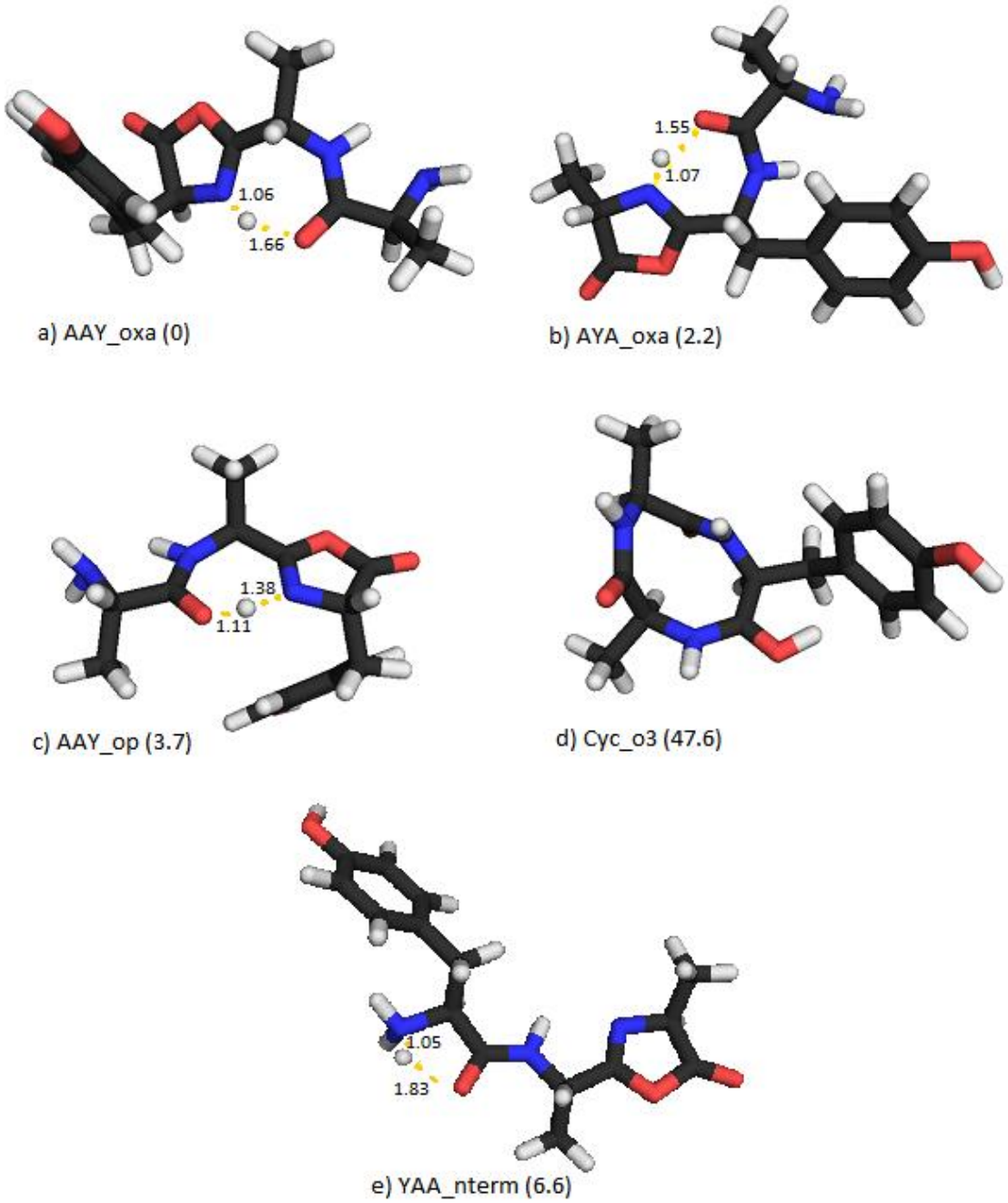

Figure 3.5. Structures of b_3^+ ions of YA_2

Lowest energy conformer is AAY_oxa among YA_2. AYA_oxa follows with 2.2 kJ/mol relative energy. Oxazolone ring protonated isomers favorable in tyrosine containing b_3^+ ions of peptide fragments. Large energy differences exist between linear and cyclic ions, 47.6 kJ/mol. YAA_nterm is the most stable N-terminally protonated conformer and has 6.6 kJ/mol energy gaps.

Phenylalanine is amino acid with aromatic group at its side chain like tyrosine and histidine. However phenyl group does not contain polar atoms or atom groups thus

shows hydrophobic properties like leucine. Table 3.6 displays to electronic and relative energies of phenylalanine containing peptide ions.

Table 3.6. Electronic Energies and zero-point corrected relative energies of FA_2

P. Site	ID	Energy(au)	ZPE(au)	Rel E. kJ/mol	Rel ZPE kJ/mol
AAF_oxa	1141	-973,4404409	-973,098535	0,0	0,0
AFA_oxa	3288	-973,4394022	-973,097964	2,7	1,5
AFA_o1	3224	-973,4385391	-973,097371	5,0	3,1
FAA_nterm	2225	-973,4399463	-973,096291	1,3	5,9
FAA_oxa	2453	-973,438321	-973,096091	5,6	6,4
FAA_o1	2368	-973,4371967	-973,095957	8,5	6,8
AAF_o1	2478	-973,4363572	-973,095708	10,7	7,4
AFA_nterm	2595	-973,4384265	-973,093748	5,3	12,6
AAF_nterm	3920	-973,4331795	-973,090024	19,1	22,3
Cyc_o3	3486	-973,4251378	-973,080422	40,2	47,6
cyc_o2	2203	-973,419063	-973,074697	56,1	62,6

Figure below shows the structure of most stable three linear and one cyclic isomer of FA_2.

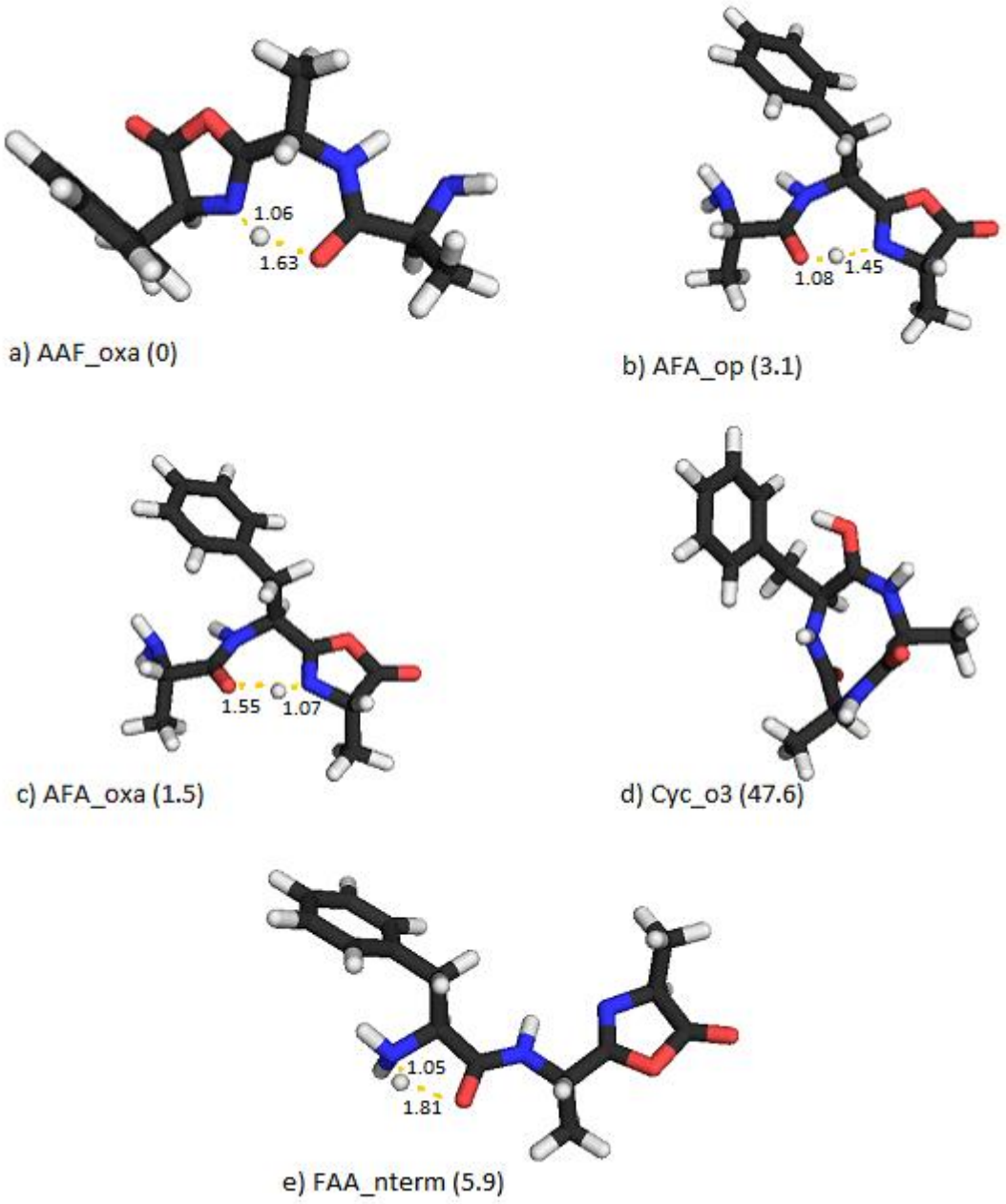

Figure 3.6. Structures of b_3^+ ions of FA_2

Energies of phenylalanine containing b_3^+ ions point out analogous trends with tyrosine. AAF_oxa is placed in global minimum of this molecule and AFA_oxa follows it with 1.5 kJ/mol relative energy (Fig3.6.a,c). Again protonation of nitrogen of oxazolone ring energetically favorable comparing to other protonation sites. Exactly the same energy barrier exists between linear and cyclic isomers of this molecule as tyrosine containing b ions. Among conformers from N-terminally protonated, FAA_nterm isomer is the most stable one and has 5.9 kJ/mol relative energy.

Cysteine is sulfur bearing amino acid. Thiol group of cysteine (-SH) have important role in the structure of proteins by creating salt bridges. In order to see how containing this group will affect structure of peptide fragments, linear and cyclic forms of b_3^+ ions created. Table 3.7 shows relative energies of CA_2.

Table 3.7. Electronic Energies and zero-point corrected relative energies of CA_2

P. Site	ID	Energy(au)	ZPE(au)	Rel E. kj/mol	Rel ZPE kj/mol
CAA_oxa	2460	-1140,556613	-1140,29613	0,0	0,0
AAC_oxa	3726	-1140,554827	-1140,29496	4,7	3,1
AAC_o1	2273	-1140,554331	-1140,29468	6,0	3,8
ACA_oxa	3584	-1140,553392	-1140,29375	8,5	6,2
ACA_o1	2797	-1140,552398	-1140,2928	11,1	8,7
CAA_nterm	2273	-1140,55422	-1140,2918	6,3	11,4
CAA_o1	1028	-1140,550439	-1140,29168	16,2	11,7
AAC_nterm	2747	-1140,552322	-1140,29074	11,3	14,2
ACA_nterm	1928	-1140,551206	-1140,28952	14,2	17,4
Cyc_o3	3756	-1140,543963	-1140,2812	33,2	39,2
Cyc_o1	3299	-1140,539426	-1140,27638	45,1	51,8
Cyc_o2	3758	-1140,535084	-1140,27238	56,5	62,3

Figure 3.7 shows the structure of most stable three linear and one cyclic isomer of CA_2.

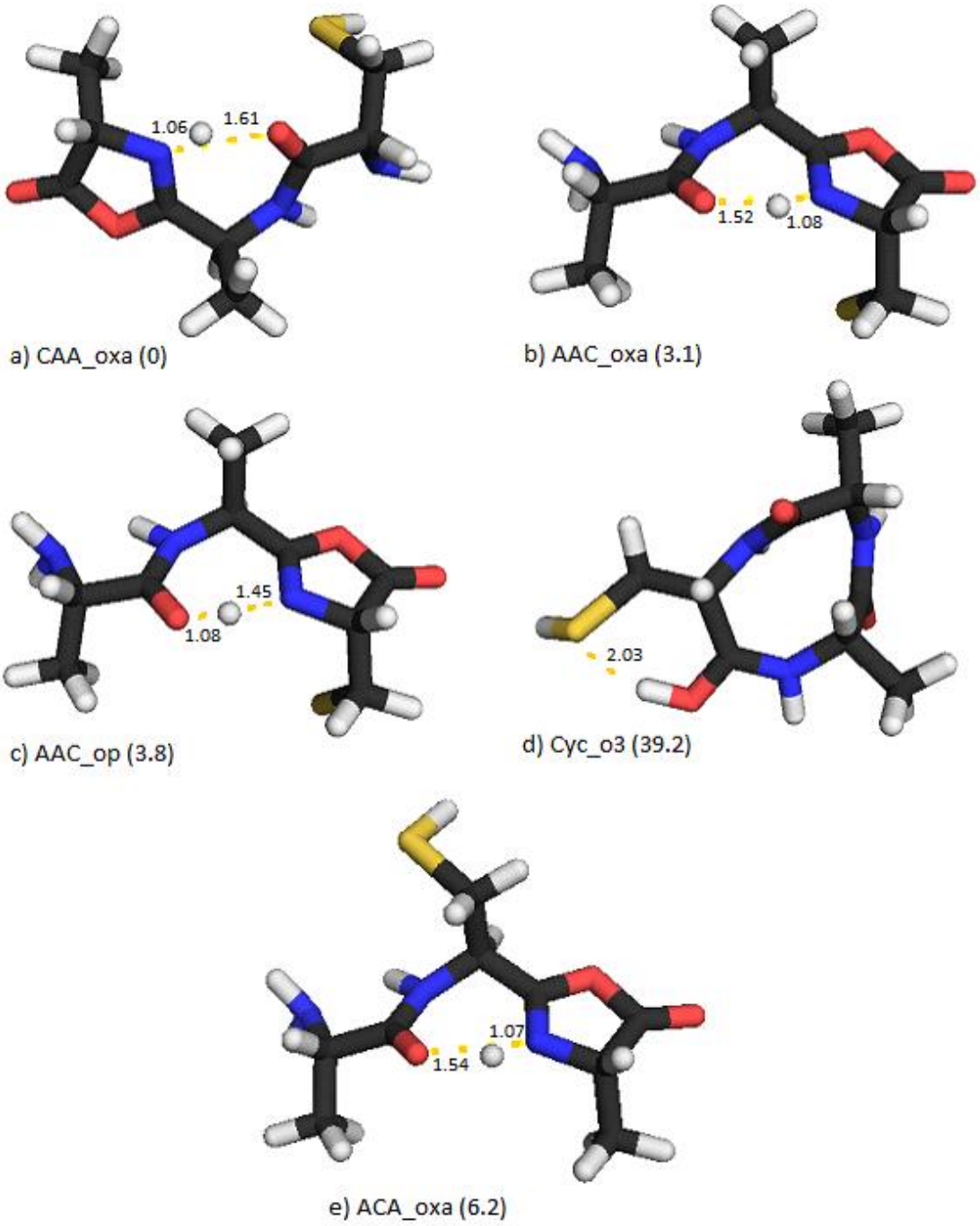

Figure 3.7. Structures of b_3^+ ions of CA_2

Oxazolone nitrogen and back bone oxygen protonated derivatives generally have lower energy than cyclic and n-terminal nitrogen protonated ones. CAA_oxa have minimum energy. After that AAC_oxa have 3.1 kJ/mol energy variations relative to minimum and AAC_o1 only differ from it as 0.7 kJ/mol. Also these two conformers are nearly same. Only difference between them is place of proton. Cyclic isomer has 39.2

kJ/mol relative energy comparing to most stable isomer and protonated from third backbone oxygen.

In order to see how configuration of the same peptide fragment ion affect its stability, relative energies of most stable isomers of b_3 ions were displayed in Figure 3.8. The energies in the figure below are given in kcal/mol. One letter code of amino acids is written in the x axis.

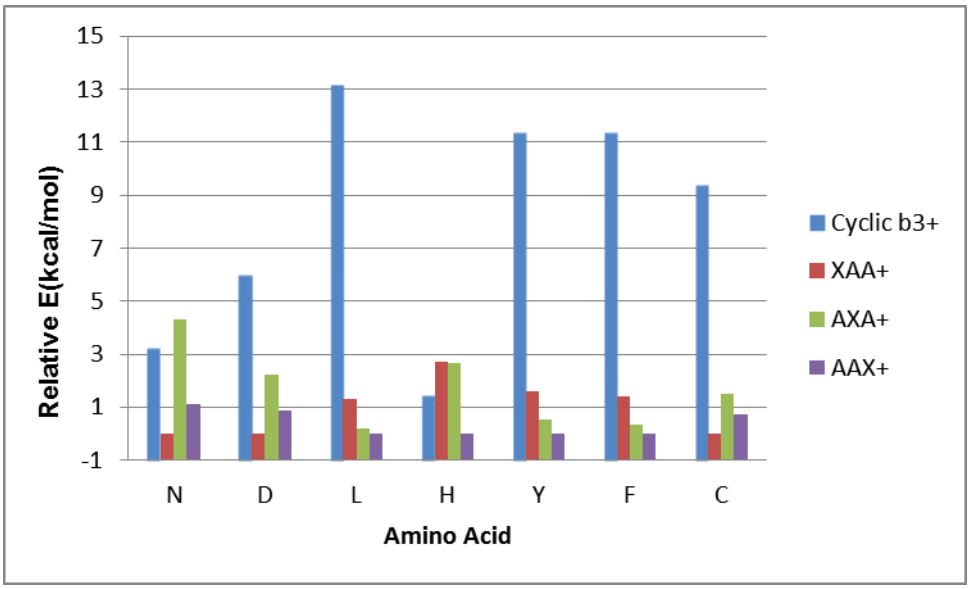

Figure 3.8. Relative energies of all constructs belongs to each amino acid

According to Figure 3.8, in five of seven amino acids cyclic positive ions are relatively high energy. Only energy of cyclic conformers from histidine and asparagine are close to the most stable conformers. However relative energy of cyclic conformer that contains aspartic acid is close to 6 kcal/mol. Leucine, tyrosine and phenyl alanine draws similar trends. Their cyclic conformers have higher than 11 kcal/mol relative energy from the most stable conformer which is in AAX configuration. Our results state that, N, D, and C residues prefers to be at N-terminal position while L, H, Y and F likes to be at C-terminal of b3-ions. It should be noted that, the energies of linear isomers are very close to other (the largest energy difference is around 1 kcal/mol) for the b_3 ions containing Y, F, C and L. Hence, there is no position effect on the energetics of those b_3 ions. For the b_3 ions having N and D, the center of the ions is less likely position as C- and N- terminals having almost equal preference. The isomers of F and Y containing b_3 ions have same relative energies (note that their structures were also very similar), that indicates the OH-group has no effect on the structure of those fragments.

3.2. b_3^o Neutral Isomers

In this section, protonated forms of all conformations of b_3^+ ions are constructed in order to find global minimum energies of neutral peptide fragments. By this way proton affinities of peptide ions can be calculated. Table 3.8 gives electrical and relative energies of neutral isomers as well as zero-point corrections.

Table 3.8. Energies with ZPE correction of neutral isomers

P. Site	Energy(au)	ZPE(au)	Rel E. kj/mol	Rel ZPE kj/mol
NA_cyc_notr	-910,7261854	-910,448146	0,0	0,0
NAA_notr	-910,718267	-910,441929	20,8	16,3
ANA_notr	-910,7149876	-910,438828	29,4	24,5
AAN_notr	-910,7132672	-910,437206	33,9	28,7
DA_cyc_notr	-930,591383487	-930,325471	0,0	0,0
DAA_notr	-930,582375772	-930,318837	23,6	17,4
ADA_notr	-930,579059927	-930,315988	32,4	24,9
AAD_notr	-930,578168460	-930,315035	34,7	27,4
LA_cyc_notr	-859,961218407	-859,625799	0,0	0,0
AAL_notr	-859,951893333	-859,619023	24,5	17,8
ALA_notr	-859,950622885	-859,617756	27,8	21,1
LAA_notr	-859,950594080	-859,617619	27,9	21,5
HA_cyc_notr	-967,048785528	-966,745859	0,0	0,0
HAA_notr	-967,040360181	-966,739339	22,1	17,1
AAH_notr	-967,040462732	-966,739281	21,9	17,3
AHA_notr	-967,037081597	-966,735693	30,7	26,7
YA_cyc_notr	-1048,298382	-1047,962517	0,0	0,0
YAA_notr	-1048,29071	-1047,957099	20,1	14,2
AAY_notr	-1048,289235	-1047,955755	24,0	17,8
AYA_notr	-1048,289115	-1047,955602	24,3	18,2
FA_cyc_notr	-973,0747253	-972,742654	0,0	0,0
FAA_notr	-973,0654822	-972,73585	24,3	17,9
AAF_notr	-973,0637194	-972,734222	28,9	22,1
AFA_notr	-973,0636191	-972,734195	29,2	22,2
CA_cyc_notr	-1140,195707	-1139,945403	0,0	0,0
CAA_notr	-1140,185639	-1139,937542	26,4	20,6
AAC_notr	-1140,184994	-1139,936674	28,1	22,9
ACA_notr	-1140,183812	-1139,935813	31,2	25,2

All of the lowest energy neutral conformers are illustrated in the figure below.

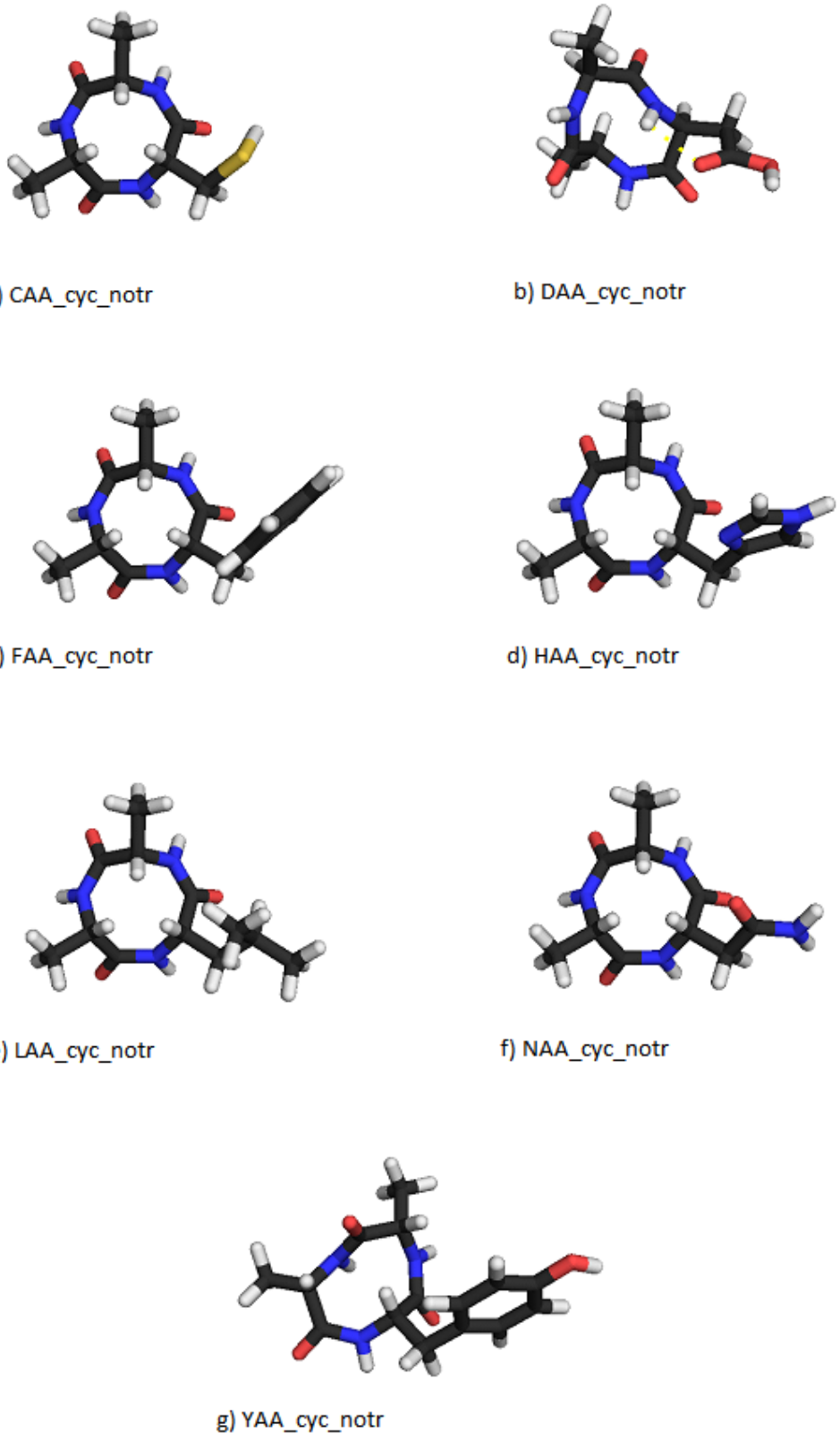

Figure 3.9. Most stable isomers of neutral peptide fragments

Different from positively charged b ions, neutral molecules tends to stay in circular formation. In all various amino acids, cyclic isomers have lowest energy. Average energy gap between cyclic and linear isomers falls into 15 to 20 kj/mol. All of the backbone peptide bonds of leucine, histidine, asparagine, phenylalanine, and cysteine containing neutral isomers are in *cis* configurations. Only tyrosine and aspartic acid contain *trans* peptide bond. Both have –OH group in their side chains. This finding may not be enough to say amino acid with –OH group containing side chain effect the conformation of neutral peptide fragments. However, it is obvious in aspartic acid inclusive cyclic neutral isomer; there is polar contact between oxygen atom in asparagine's side chain and back bone nitrogen which is indicated as yellow dots in Figure 3.9.b. Second stable isomers of all peptide ion configurations are XAA, which X refer to one letter code of different amino acids; N, D, H, F, Y, C. Within linear conformers, only leucine containing ones have AAX configuration.

3.3. Proton Affinity

Proton affinity can only be defined in gas phase of atoms or molecules. It is described as negative enthalpy change during the reaction of negative or neutral atoms or molecules that take hydrogen atom. Formula below is indicating the reaction.

$$A_{(g)} + H^{+}_{(g)} \rightarrow AH^{+}_{(g)} \qquad (3.1)$$

Normally, basic amino acids residues expected to have higher proton affinities. Since probability of the abundance of peptide fragment ions and relative intensities in peptide mass spectra depend on proton affinity of the ions, knowing proton affinities have vital importance to understand rules behind these phenomena.

So far, positive b_3 ions and neutral isomers of various amino acid compositions have been investigated. Energies and three dimensional structures elucidated. Using this information, we calculate proton affinities of each b_3 ions molecules. Table below show lowest zero-point corrected energies of positive and neutral constructs and their corresponding proton affinities. These proton affinities were calculated based on the most stable positive and neutral conformers without taking into account of their sequences (that is the most stable forms of XA_2^{+} ions).

Table 3.9. Proton affinities of various b ions

AA	$E_{Positive}$(au)		$E_{Neutral}$(au)		ΔE(kcal/mol)
Cys	**CAA_oxa**	-1140,296129	**CA_cyc_notr**	-1139,9454	**220,1**
Asp	**DAA_oxa**	-930,679694	**DA_cyc_notr**	-930,325471	**222,3**
Phe	**AAF_oxa**	-973,098535	**FA_cyc_notr**	-972,742654	**223,3**
Leu	**AAL_oxa**	-859,981795	**LA_cyc_notr**	-859,625799	**223,4**
Tyr	**AAY_oxa**	-1048,321232	**YA_cyc_notr**	-1047,96252	**225,1**
Asn	**NAA_oxa**	-910,812209	**NA_cyc_notr**	-910,448146	**228,5**
His	**AAH_his**	-967,127212	**HA_cyc_notr**	-966,745859	**239,3**

"AA" part of the table indicate which amino acid involve in b ions, "$E_{positive}$" and "$E_{neutral}$" shows energy of positively charged and neutral conformers and their protonation site and compositions. ΔE is energy difference between positive and neutral isomers of the same peptide ion which indicate proton affinity of this molecule. Unit of the energy used to indicate proton affinities are kcal/mol. Graphical representation of proton affinities are demonstrated below with additional information of experimental proton affinities from the literature (Fig. 3.9) (A. G. Harrison, 1997).

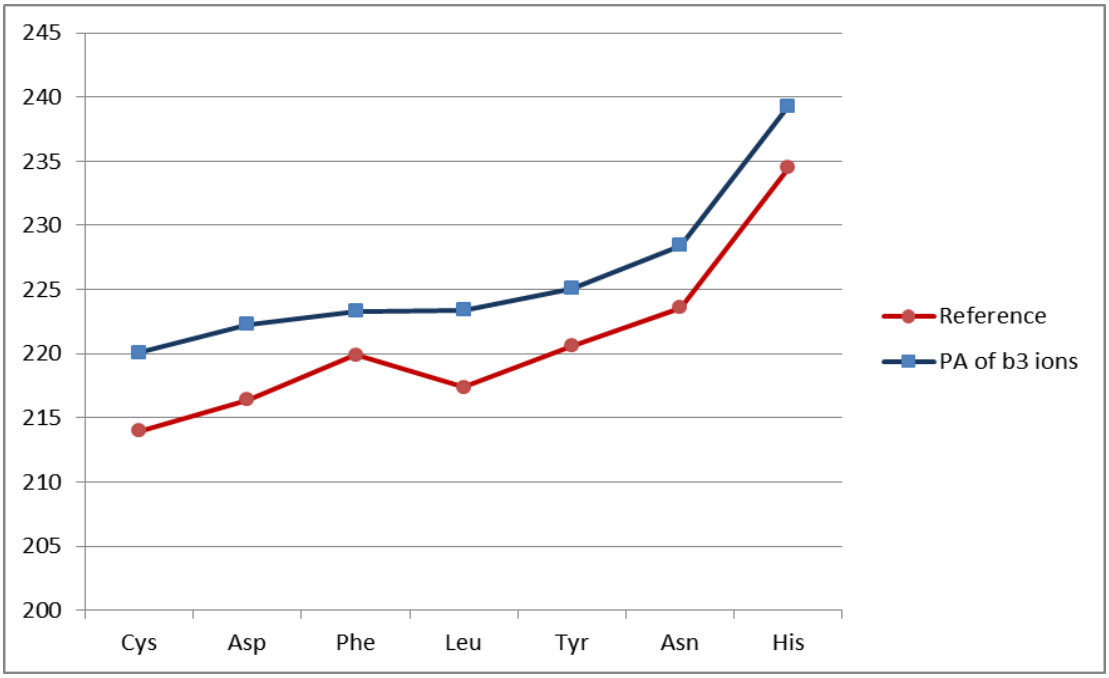

Figure 3.10. Graphical representation of proton affinities from literature reference and this study. Blue line indicates values calculated in this thesis. Red line indicates values calculated in literature reference (A. G. Harrison, 1997). In the reference, PA's belongs to single amino acids.

Values from literature belong to single amino acids and calculation from this study belongs to three amino acid long peptide ions. However, except phenylalanine similar trend is observed between both proton affinity calculations. As expected and parallel to literature work histidine containing b_3 ion have higher proton affinity (240 kcal/mol). Asparagine follows histidine with 228.5 kcal/mol proton affinity energy, but between them there are relatively big energy differences comparing to others. Cysteine containing b_3 ions has lowest proton affinity among investigated amino acids.

Additionally, various affinity calculations were made in order to see effect of structure on proton affinity. In this time, energy of linear or cyclic b_3^+ ions compared to their linear or cyclic neutral forms. This calculation gives us comparison of proton affinities from both cyclic and linear isomers which gives better viewpoint to look proton affinities more deeply. Following table summarize the information used in this purpose.

Table 3.10. Proton affinities of cyclic and linear conformers

	PA(cyclic)	PA(lineer)	PA	Δ
Cys	210,7	225,0	220,1	14,3
Asp	216,3	226,4	222,3	10,2
Phe	212,0	227,6	223,3	15,6
Leu	210,2	227,6	223,4	17,4
Tyr	213,7	228,5	225,1	14,8
Asn	225,2	232,4	228,5	7,1
His	237,8	243,4	239,3	5,5

To find proton affinity of cyclic peptides, energy of the most stable structure of cyclic b_3^+ ions and neutral forms are used. The same procedure is done for proton affinity of linear conformers. PA values come from Table 3.9. Delta represents proton affinity differences of linear and cyclic conformers. All of the values in the table are in kcal/mol. Following figure shows the distribution of both cyclic and linear PA values as well as PA values from previous calculation. Three letter codes of amino acids are in the x axis of the figure.

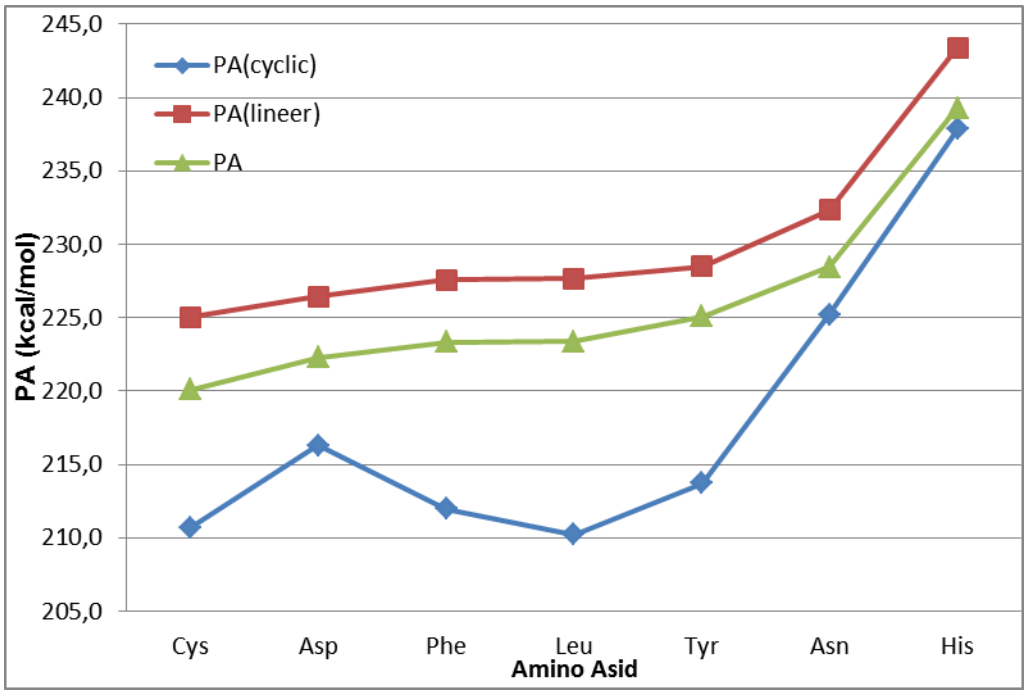

Figure 3.11. Proton affinity values of linear and cyclic conformers. Y axis indicates energy in kcal/mol. Color codes of lines represent PA values of corresponding structures.

Proton affinity of linear structures is relatively higher than cyclic structures in all peptide fragments. However, the difference between cyclic and linear PA values is not the same for all amino acids. For example, histidine has smaller PA difference between its cyclic and linear conformers. Also trend of PA values from cyclic and linear peptides are not same. PA value of cyclic peptide which contain aspartic acid have significantly higher than PA of cyclic Cys, Phe and Leu. Following graph shows the differences between cyclic and linear structures of each amino acid.

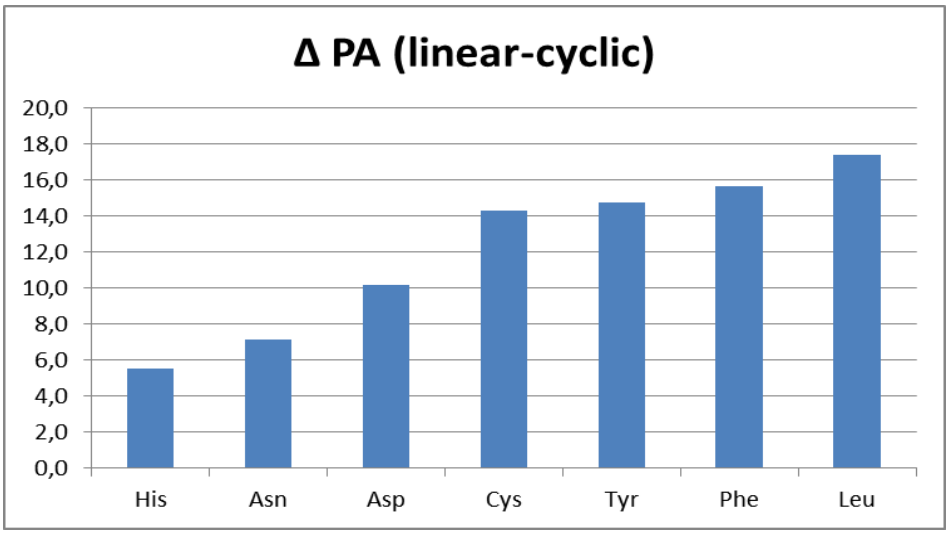

Figure 3.12. Differences of PA values in linear and cyclic conformers. X axis shows the difference of PA values between cyclic and linear conformers for each peptide ions in kcal/mol. Name of amino acid given in y axis.

Largest PA difference between cyclic and linear isomers exists in leucine containing peptide fragments which is almost 18 kcal/mol. Histidine and asparagine containing peptides have smaller PA value difference between their cyclic and linear conformers. Aspartic acid has about 10 kcal/mol difference.

In order to see is there any effect of the location of specific amino acid among peptide fragment to proton affinity; we calculate PA values of each cyclic XAA, linear XAA, AXA, and AAX peptides. Following table shows each PA values corresponding to their appropriate peptide fragments.

Table 3.11. Relative PA values of various peptide fragments

Name	**Phenyl**		**Histidine**		**Cysteine**		**Tyrosine**	
	H^+ Site	PA	H^+ Site	PA	H^+ Site	PA	H^+ Site	PA
XAA_cyc	o3	0,0	his	0,0	asn_op	0,0	o3	0,0
XAA	nterm	14,2	his	2,8	oxa	14,3	nterm	13,2
AXA	oxa	16,3	his	5,2	oxa	13,9	oxa	15,2
AAX	nterm	16,7	his	5,6	oxa	14,1	oxa	15,6
Name	**Leucine**		**Asparagine**		**Aspartic acid**			
	H^+ Site	PA	H^+ Site	PA	H^+ Site	PA		
XAA_cyc	o1	0,0	asn_op	0,0	o1	0,0		
XAA	oxa	17,0	oxa	7,1	oxa	10,2		
AXA	oxa	18,0	oxa	4,8	oxa	9,7		
AAX	oxa	17,4	nterm	9,0	nterm	11,7		

Name indicates the structure of peptide fragments. Name of the various amino acids written in bold and two adjacent columns belong to this amino acid. The row labeled with proton site shows which protonation site is most stable for that structure. PA values are relative and be viewed as kcal/mol. Clearly, all of the cyclic structures have lower proton affinity than linear conformers. However, in some amino acids proton affinity difference between cyclic and linear conformers are very low. Especially, proton affinities of cyclic and linear histidine very close. Only 2.8 kcal/mol energy differences exist between them. Also, similar feature is observed for proton affinities of cyclic and linear peptides which contain asparagine and aspartic. Generally, linear conformers whose specific amino acids close to carboxylic end have higher proton affinity.

CHAPTER 4

CONCLUSION

In this thesis, structures and electrochemical properties of various amino acids and their different compositions on b_3+ ions and their neutral states were investigated. In addition, proton affinities of each b ions construct mentioned. The most obvious deduction from the results of this study, b_3^+ ions prefers to have linear oxazolone structure. However, in their neutral states this is not observed, cyclic structures are relatively far more stable than linear isomers. In neutral conformers, energy differences can be varied between 16.3 kJ/mol and 20.6 kJ/mol. On the other hand, in positive ions this difference is much larger, beginning from 6.1 kJ/mol and up to 55.1 kJ/mol. Histidine containing ions have tiny energy gap between cyclic and linear conformers. Also asparagine and aspartic acid containing ions relatively have smaller energy differences between linear and cyclic ions (around 15 and 25 kJ/mol respectively). Leucine have largest energy gap among whole amino acid used in this study. According to these findings b_3^+ ions which contain amino acids with basic and polar side chains generally have smaller energy differences between linear and cyclic isomers comparing to aliphatic and aromatic side chain containing ones.

The stable isomers of b_3^+ have different position of amino acid based on their chemical properties. The proton and the side chain position of three most stable isomers of both asparagine and aspartic acid are the same. These amino acids prefer to stay close to N-terminal, and protonated from oxazolone ring. The second stable isomers are N-terminal protonated ones and their side chains close to C-terminal in both of these amino acids. Additionally leucine, tyrosine and phenylalanine shows similar trend comparing the position of their side chains along stable peptide ions. Different from asparagine and aspartic acid their side chain form more stable structure when they are close to C-terminal. Also N-terminal protonation site is energetically less favorable comparing to polar amino acids. Histidine display different behavior than other amino acids. Side chain of histidine holds protons and forms stable structures. Hydrogen bonding occurs between protonated histidine side chain and back bone oxygen in the most stable histidine isomers. The energies of cyclic and linear isomers of His containing b ions are close to each other. Cysteine side chain has lower energy when it

closes to N-terminal like asparagine and aspartic acid. On the other hand oxazolone protonated conformers have relatively stable comparing to N-terminally protonated ones. Cyclic structures have much higher energies relative to most stable conformers except histidine and asparagine.

In this study, we calculate proton affinities of eight different peptide fragments. The only difference between each other is one specific amino acid. Thus, we manage to observe the effect of this specific amino acid into proton affinity of corresponding peptide fragments. Also, calculating the proton affinity of linear and cyclic conformers with various compositions enable us to see is there any change in terms of proton affinity when we use different conformers of same molecule. According to our findings, histidine containing peptide fragments have larger proton affinity comparing to others. Probably, basic side chain of histidine causes this result. Comparing to experimental results from literature, our calculations draw similar trends considering order of proton affinities. Difference of proton affinities between linear and cyclic conformers varies based on amino acid used. This difference is lower than 10kcal/mol in histidine, asparagine and aspartic acid containing peptide fragments.

This thesis study revealed valuable information and insights about geometry, structure and chemical properties of b ions and hopefully will help further studies to develop more efficient algorithms for peptide identification.

REFERENCES

Atik, a E., & Yalcin, T. (2011). A systematic study of acidic peptides for b-type sequence scrambling. *Journal of the American Society for Mass Spectrometry*, *22*(1), 38–48. http://doi.org/10.1007/s13361-010-0018-3

Becker, O. M. (1998). Potential Energy Surfaces, *19*(11), 1255–1267.

Bleiholder, C., Osburn, S., Williams, T. D., Suhai, S., Van Stipdonk, M., Harrison, A. G., & Paizs, B. (2008). Sequence-scrambling fragmentation pathways of protonated peptides. *Journal of the American Chemical Society*, *130*(52), 17774–89. http://doi.org/10.1021/ja805074d

Bleiholder, C., Suhai, S., Harrison, A. G., & Paizs, B. (2011). Towards understanding the tandem mass spectra of protonated oligopeptides. 2: The proline effect in collision-induced dissociation of protonated Ala-Ala-Xxx-Pro-Ala (Xxx = Ala, Ser, Leu, Val, Phe, and Trp). *Journal of the American Society for Mass Spectrometry*, *22*(6), 1032–9. http://doi.org/10.1007/s13361-011-0092-1

Case, D. A., Darden, T. A., Cheatham, T. E., Simmerling, C. L., Wang, J., Duke, R. E., … Kollman, P. A. (2011). Amber 11.

Chass, G. A., Marai, C. N. J., Setiadi, D. H., Csizmadia, I. G., & Harrison, A. G. (2004). A Hartree – Fock , MP2 and DFT computational study of the structures and energies of 00 b 2 ions derived from deprotonated peptides . A comparison of method and basis set used on relative product stabilities q, *675*, 149–162. http://doi.org/10.1016/S0166-1280(03)01065-0.

Chen, X., & Turecek, F. (2005). Simple b ions have cyclic oxazolone structures. A neutralization-reionization mass spectrometric and computational study of oxazolone radicals. *Journal of the American Society for Mass Spectrometry*, *16*(12), 1941–56. http://doi.org/10.1016/j.jasms.2005.07.023

Dunteman, G. H. (1989). *Principal Components Analysis.* Sage Publications. Retrieved from http://libezproxy.iyte.edu.tr:81/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=24769&site=eds-live

Elmaci, N., & Berry, R. S. (1999). Principal coordinate analysis on a protein model. *The Journal of Chemical Physics*, *110*(21), 10606. http://doi.org/10.1063/1.478992

Engler, E. M., Andose, J. D., & Schleyer, P. V. R. (1973). Critical Evaluation of Molecular Mechanics. *J. Am. Chem. Soc.*, *95*, 8005.

Erlekam, U., Bythell, B. J., Scuderi, D., Van Stipdonk, M., Paizs, B., & Maître, P. (2009). Infrared spectroscopy of fragments of protonated peptides: direct evidence for macrocyclic structures of b5 ions. *Journal of the American Chemical Society*, *131*(32), 11503–8. http://doi.org/10.1021/ja903390r

Foresman, J. B., & Frisch, A. (1996). Exploring Chemistry With Electronic Structure Methods.pdf. *Exploring Chemistry with Electronic Structure Methods*.

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., … Fox, D. J. (2009). Gaussian 09 Revision D.01. Gaussian, Inc., Wallingford CT.

Fu, L., Chen, T., Xue, G., Zu, L., & Fang, W. (2013). Selective cleavage enhanced by acetylating the side chain of lysine. *Journal of Mass Spectrometry : JMS*, *48*(1), 128–34. http://doi.org/10.1002/jms.3136

Grebner, C., Becker, J., Stepanenko, S., & Engels, B. (2011). Efficiency of tabu-search-based conformational search algorithms. *Journal of Computational Chemistry*, *32*(10), 2245–53. http://doi.org/10.1002/jcc.21807

Gu, C., Tsaprailis, G., Breci, L., & Wysocki, V. H. (2000). Selective Gas-Phase Cleavage at the Peptide Bond C-Terminal to Aspartic Acid in Fixed-Charge Derivatives of Asp-Containing Peptides, (10), 5804–5813.

Harrison, A. (2009). To b or not to b: the ongoing saga of peptide b ions. *Mass Spectrometry Reviews*, (November 2008), 640–654. http://doi.org/10.1002/mas

Harrison, A. G. (1997). the Gas-Phase Basicities and Proton Affinities of Amino Acids and Peptides. *Mass Spectrometry Reviews*, *16*, 201–217. http://doi.org/10.1002/(SICI)1098-2787(1997)16:4<201::AID-MAS3>3.0.CO;2-L

Harrison, A. G. (2009a). Fragmentation reactions of some peptide b 3 ions : an energy-resolved study, 1298–1302. http://doi.org/10.1002/rcm

Harrison, A. G. (2009b). TO b OR NOT TO b : THE ONGOING SAGA OF PEPTIDE b IONS, (November 2008), 640–654. http://doi.org/10.1002/mas

Harrison, A. G., Young, A. B., Bleiholder, C., Suhai, S., & Paizs, B. (2006). Scrambling of sequence information in collision-induced dissociation of peptides. *Journal of the American Chemical Society*, *128*(32), 10364–5. http://doi.org/10.1021/ja062440h

Jolliffe, I. T. (2002). *Principal component analysis*. New York : Springer, 2002. Retrieved from http://libezproxy.iyte.edu.tr:81/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=cat01635a&AN=iyte.212129&site=eds-live

Li, J. (2005). 2.8 Basic Molecular Dynamics, 565–588.

Molesworth, S., Osburn, S., & Van Stipdonk, M. (2010). Influence of amino acid side chains on apparent selective opening of cyclic b5 ions. *Journal of the American Society for Mass Spectrometry*, *21*(6), 1028–36. http://doi.org/10.1016/j.jasms.2010.02.011

Orio, M., Pantazis, D. a, & Neese, F. (1983). Density functional theory. *Photosynthesis Research*, *102*(2-3), 443–453. http://doi.org/10.1007/s11120-009-9404-8

Paizs, B., & Suhai, S. (2005). Fragmentation pathways of protonated peptides. *Mass Spectrometry Reviews*, *24*(4), 508–48. http://doi.org/10.1002/mas.20024

Paizs, B., Suhai, S., Hargittai, B., Hruby, V. J., & Somogyi, Á. (2002). Ab initio and MS / MS studies on protonated peptides containing basic and acidic amino acid residues I . Solvated proton vs . salt-bridged structures and the cleavage of the terminal amide bond of protonated RD-NH 2, *219*, 203–232.

Rapaport, D. C. (1997). The Art of Molecular Dynamics Simulation. *Book*.

Rapaport, D. C. (1999). Molecular Dynamics Simulation. *Computing in Science and Engineering*, *1*(1), 70–71.

Rappé, A. K., & Casewit, C. J. (1997). Molecular Mechanics Across Chemistry. Retrieved from http://books.google.com/books?id=U0TeILEDTr0C&pgis=1

Sun, F., Liu, R., Zong, W., Tian, Y., Wang, M., & Zhang, P. (2010). A unique approach to the mobile proton model: influence of charge distribution on peptide fragmentation. *The Journal of Physical Chemistry. B*, *114*(19), 6350–3. http://doi.org/10.1021/jp911772q

Tang, T.-H., Fang, D.-C., Harrison, A. G., & Csizmadia, I. G. (2004). A computational study of the fragmentation of b3 ions derived from protonated peptides. *Journal of Molecular Structure: THEOCHEM*, *675*(1-3), 79–93. http://doi.org/10.1016/j.theochem.2003.12.033

Xu, C., & Ma, B. (2006). Software for computational peptide identification from MS-MS data. *Drug Discovery Today*, *11*(13-14), 595–600. http://doi.org/10.1016/j.drudis.2006.05.011

Young, D. (2004). *Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems*. Wiley. Retrieved from https://books.google.com.tr/books?id=-pn8K53IUqgC