DEMOCRITUS UNIVERSITY OF THRACE
HEALTH SCIENCES - SCHOOL
DEPARTMENT OF MOLECULAR BIOLOGY & GENETICS

BSc Thesis

Validation of the AMBER-15SB force field using an α-Lactalbumin-derived peptide

Author:
Buse Kaymaz
Dept. of Molecular Biology and Genetics
Faculty of Science and Letters
Istanbul Technical University

Supervisor:
Dr. Nicholas M. Glykos
Assistant Professor of Structural and Computational Biology
Department of Molecular Biology and Genetics
Health – Sciences School
Democritus University of Thrace

Alexandroupolis, Greece, June 2019

ISTANBUL TECHNICAL UNIVERSITY ★ FACULTY OF SCIENCE AND LETTERS

VALIDATION OF THE AMBER-15SB FORCE FIELD USING AN α-LACTALBUMIN-DERIVED PEPTIDE

B.Sc THESIS

BUSE KAYMAZ

(090140039)

Department of Molecular Biology and Genetics

Thesis Advisor: Dr. Nicholas M. Glykos

JUNE 2019

FOREWORD

This thesis was written under the supervision of Assoc. Prof. Nicholas M. Glykos at Democritus University of Thrace, Greece, where I came with Erasmus Learning Program.

First of all, I would like to extend my sincere gratitude to Assoc. Prof. Nicholas M. Glykos for giving me the opportunity to work with him, showing his help and guidance throughout my work, and sharing his knowledge and experience with me.

Also, I am very thankful to my Erasmus advisor Assoc. Prof. Aslı Kumbasar for helping me provide this opportunity and being an exemplary instructor.

I place on record, many thanks to my dear friends Peri Sherifli, Ahmet Fatih Ayaş, Cemil Can Saylan, and Murat Doğu for being always in my side as well as Aycan Tül who lives in Greece for all kinds of support and friendship. In addition, I want to thank to Ceren Sağdıç, PhD candidate, one of my supporters in my life for her guidance, kindness, and friendship.

Lastly, I would like to thank my dear family for supporting and encouraging me.

June 2019 Buse Kaymaz

TABLE OF CONTENTS

ABBREVIATIONS

3D : Three dimensional

4fs-STAR : AMBER99SB-4fs-STAR-ILDN

15SB : AMBER15SB

Atm : Atmosphere

dPCA : Dihedral Angle Principal Component Analysis

MD : Molecular Dynamics

NMR : Nuclear Magnetic Resonance

NOE : Nuclear Overhauser Effect

RMSD : Root Mean Square Deviation

STAR : AMBER99SB-STAR-ILDN

αLa : α-Lactalbumin

Å : Angstrom

μs : Microsecond

K° : Kelvin

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Molecular Dynamics is widely used to define their structures and to study the folding process by simulating the motion of molecules. In this thesis, validation of AMBER15SB force field is aimed by using INYWLAHAKAG peptide derived from the residues 101-111 of α-lactalbumin. Trajectory analyzes were performed with previously performed simulations of the three force fields, comparing the different force fields such as AMBER99SB-STAR-ILDN and AMBER99SB-4fs-STAR-ILDN with experimental data. RMSD analysis showed the stability of the peptide, while the secondary structure analysis showed the possibility of secondary structure that could occur during the simulation time of each residue of the peptide. In the dPCA analysis, it created two-dimensional landscapes and provided us with information about their energy minima. NOE, one of the observable NMR commonly used when comparing experimental data, is related to the interatomic distance of the two nuclei affecting each other. It was determined whether NOE values had upper bound violation according to the experimental data and averages of violations were found to be close to experimental data. As a result, several trajectory analyzes were performed within the three force fields. The validation of the AMBER15SB force field was performed and the average of violations according to NOE values was found to be 0.052 A, which is a reasonable value. AMBER99SB-STAR-ILDN was found to have the closest results to the experimental data with an average violation value of 0.0.46 A.

ÖZET

Moleküler Dinamik, moleküllerin hareketlerini simüle ederek onların yapılarının tanımlanmasında ve katlanma sürecinin incelenmesinde yaygın olarak kullanılır. Bu tezde α-laktalbumin'in 101-111 kalıntılarından türetilmiş INYWLAHAKAG peptidi kullanılarak AMBER15SB kuvvet alanının validasyonu amaçlanır. Sistemin izlediği yol analizleri, AMBER99SB-STAR-ILDN ve AMBER99SB-4fs-STAR-ILDN gibi farklı kuvvet alanlarının deneysel verilerle karşılaştırılarak üç kuvvet alanının önceden gerçekleştirilmiş simülasyonlarıyla yapıldı. Sistemin izlediği yol analizlerinden RMSD analizi peptidin kararlılığını gösterirken ikincil yapı analizi peptidin her kalıntısının simülasyon zamanı boyunca oluşabilecek ikincil yapı olasılıklarını gösterdi. dPCA analizinde iki boyutlu enerji yüzeyleri oluşturarak onların düşük serbest enerjileri hakkında bilgi sahibi olmamızı sağladı. Deneysel veriyle karşılaştırma yapılırken yaygın olarak kullanılan gözlemlenebilir NMR'dan bir olan NOE, birbirini etkileyen iki çekirdeğin atomlar arası uzaklığıyla ilişkilidir. NOE değerlerinin deneysel veriye göre yüksek sınır ihlalleri hesaplandı ve bu ihlallerin ortalamaları alınarak deneysel veriye yakınlığı saptanmıştır. Sonuç olarak, üç kuvvet alanı içinde birçok sistemin izlediği yol analizi yapıldı. AMBER15SB kuvvet alanının validasyonu gerçekleştirildi ve NOE değerlerine göre ihlallerin ortalaması makul bir değer olan 0.052 A olarak bulunmuştur. Üç kuvvet alanı içinde AMBER99SB-STAR-ILDN 0.0.46 A ortalama ihlal değeri ile deneysel veriye en yakın sonuclara sahip olduğu saptanmıştır.

1. INTRODUCTION

1.1. Proteins

Proteins are biomolecules, specially polypeptides, that formed from one or more chains of amino acid residues. Proteins take on various tasks in reactions of organisms for example metabolic reactions, DNA replication, transcription-translation and transporting molecules, etc., is essential for life. [1]

Each protein has got different amino acid sequence and folding shapes that's why they are unique. Polypeptides consist of amino acids that are bonded as linear to each other with peptide bonds which form by bonding from two different end of the amino acids like $–NH_2$ amino group and –COOH carboxyl group. This linear structure represents the primary structure of the protein. As a result of the approach of amino acid residues in a certain ratio, hydrogen bonds are formed between C=O group and N-H group of different peptide bond. The regular and local sub-structures formed consist of two structures, α-helix and β-sheet, which form secondary structures. A 3-dimensional compact globular structure forms by folding with strong interaction of amino acids belong to a polypeptide. The tertiary structure known as the natural state of the folding protein is formed. Protein has got functionality with processing of protein folding. Quaternary structure is a compact structure composed of multiple polypeptides interacting with each other. [1]

According to the Anfinsen hypothesis, the native structure for a globular protein, which is thought to be present in the standard physiological environment, depends on its amino acid sequence. In addition, the minimum of free energy of a protein must supply three conditions for its native structure: uniqueness, stability, and kinetical accessibility. [2] To understand how a protein is formed, the protein folding subject associated with Levinthal's paradox should be addressed.

1.2. Protein folding

Protein folding is a quick and repeatable physical process which obtaining a native 3D structure that is biologically functional from a protein chain. Each protein is found to be unstable in random coils or not properly folded when converted from the mRNA sequence to the amino acid chain. As the translation continues on the ribosome, folding starts on the linear amino acid chain. In order to obtain a 3-dimensional protein with a stable structure, the amino acids must interact with each other. [3] This is confirmed by the fact that the 3-dimensional structure of the

protein is determined its primary structure according to the Anfinsen hypothesis. [4] Some parts of proteins with the correct 3-dimensional structure may remain in the unfolded state but are still functional. In this case, protein dynamics can be mentioned.

The duration of the protein folding process may vary depending on the protein of interest. In studies conducted outside the cell, the slowest folding proteins due to primarily the presence of proline isomerization are performed for several minutes or, hours, [5] while protein folding of single domain proteins that can be folded in one step can take place in several microseconds. [6]

1.2.1. Computational studies of protein folding

Computational studies of protein folding shed light on the three main directions of protein stabilization, kinetics and structure.

1.2.1.1. Levinthal's paradox

In 1969, Cyrus Levinthal said that an unfolded polypeptide chain had an astronomical number of conformations based on it had too many degrees of freedom. According to estimates, this number is approximately 3^{300} or 10^{143}. This paradox is a thought experiment, which is assumed to take a considerable amount of time if all the possible conformations of a protein are sampled respectively even if each is assumed to be nanosecond. Furthermore, Levinthal indicated that a random conformation of the protein cannot be selected if proteins are considered to be folded faster than the assumed. Instead, a number of proteins can be observed in the meta-stable intermediate state. [7]

1.2.1.2. Energy landscape – Folding funnels

The configuration possibilities that the protein may generate during folding can be visualized as the energy field. Joseph Bryngelson and Peter Wolynes explain the acquisition of functionalized folded protein by the principle of minimal frustration as a result of the optimization of energy landscapes of all naturally evolved proteins with topological and energy barriers. The amino acid sequences evolve to make the folded protein sufficiently stable. [8] As a result of this evolution, the proteins are thought to have a globally funneled energy landscapes as shown

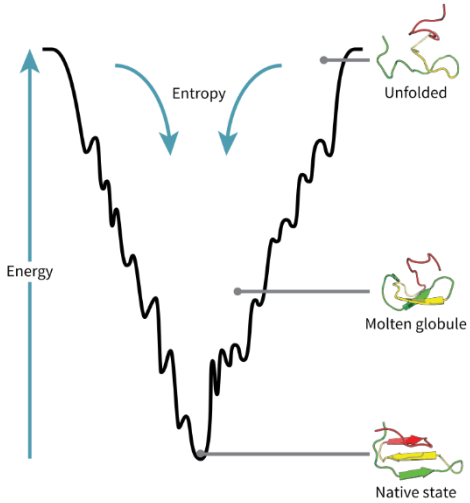


Figure 1: A funneled energy landscape showed how unfolded polypeptide fold into their native structures by minimizing their free energy. [adapted without permission from *Wikipedia*.]

in Figure 1 which produced by José Onuchic. This landscape ensures nature state of protein by way of many pathways and intermediate instead of limiting a mechanism to them. Physically, thinking of landscapes with potential and total energy can be more complex and misleading with maxima, saddle points, minima, and funnels rather than geographic concepts. [9]

The theory is supported by experimental and computational studies and is also consistent with the 2nd law of thermodynamics. Using Entropy and Enthalpy, the energy difference between the folded and unfolded states of a protein can be calculated. According to the 2nd law of thermodynamics, the higher the entropy, the more favourable unfolded protein. As the enthalpy increases, the folded protein is favourable. As the irregularity decreases while folding the protein native state, entropy is reduced. [10] At the same time, interactions such as hydrogen bonds and van der Waals interactions that help maintain the protein's stable state begin to occur. Entropy and Enthalpy relationship is explained by Gibbs free energy formula.

$$\Delta G = \Delta H - T\Delta S$$

According to the magnitude of Free-energy, it is understandable whether the protein will be folded or not. Gibbs Free Energy is closely related to folding funnel. When Gibbs free energy reaches the minimum magnitude (negative), the protein is considered to be close to the native state and placed in the bottom of the funnel. Landscapes consist of local minima in various locations related to non-native proteins. Thermodynamically non-stable proteins with high energy can be placed in the local minima in the upper parts of the folding funnel. [11]

1.2.1.3. Modeling of Protein Folding

Using the de novo and ab initio techniques, the folding of the protein of interest can be predicted computationally. Molecular Dynamics (MD) has been a tool in research on protein folding and its dynamics in the silico medium. [12] First equilibrium folding simulation was prepared using implicit solvent model and umbrella sampling. Explicit solvent model is limited to very small proteins due to simulation costs. The Markov state model, which is assumed to be a Markov process, can be used account for cases that cannot be observed in the system in order to determine the fold pathways by modeling a protein from the randomly ordered sequence to the native 3-dimensional structure. [13]

Hydrophobic collapse hypothesis is based on the fact that the spherical proteins have the hydrophobic core consists of their non-polar amino acids while their polar amino acids are placed outside of them interact with water molecules. According to hypothesis, first secondary

structures form due to hydrophobic interactions and so intermediate form which has free energy is less than free energy of unfolded state and more than native folded state is formed. [14] In addition, the Nucleation-Condensation mechanism tends to be nucleated by the effect of hydrophobic interactions due to the formation of secondary and tertiary structures at the same time. The reason why there is no obvious energy differences between the unfolded and folded states of the proteins is that each protein has its own unique structures, which are adopted as domains. Increasing the stability of the structures to become more condense by interacting them with each other is the basic principle of this system. [15] Another hypothesis, Diffusion-Collision hypothesis, was reported by Martin Karplus and David L. Weaver, in which proteins are composed of small micro-domains, these structures are diffuse and collide with each other, and these structures guide folding estimates since all possible doubts are unknown. [16]

1.2.1.4. Experimental techniques for studying protein folding

When studying protein folding with mutation studies, inferences can be made by using standard non-crystallographic techniques to observe conformational changes and how it gradually folded using experimental techniques. X-ray crystallography is one of the effective methods to define the three-dimensional structure of a folded protein. The secondary and tertiary structures can be identified by inserting the interest protein into the crystal lattice in the most suitable solution to form well-defined crystals when X-ray diffraction occurs. [17] Fluorescence spectroscopy is one of the sensitive methods for studying the folding of the protein by utilizing the specific fluorescence properties of amino acids such as phenylalanine (Phe), tyrosine (Tyr) and tryptophan (Trp). [18] Nuclear magnetic resonance (NMR) spectroscopy, for example by monitoring hydrogen-deuterium exchange, is a common technique used to inform about the structure and dynamics of protein. [19] Circular dichroism spectroscopy is one of the most basic tools to examine protein folding by measuring the absorption of circular polarized light. Structures as α helices and β sheets are found in proteins have chiral so they absorb the light. [20] Instruments such as mass spectroscopy, AFM (atomic force microscopy), FTIR spectroscopy (Fourier Transform Infrared Spectroscopy) also shed light on protein folding studies.

All of the techniques mentioned above are about estimation of the structure, further studies are possible with computational studies. The Molecular Dynamic Simulations, which is the definitive calculation tool, try to define the protein folding and its structure by linking the experiments and theories.

1.3. Molecular Dynamics Simulation

Molecular Dynamics (MD) are the science of simulating the physical motions of atoms and molecules. Although the method was first introduced at the end of the 1950s in the field of theoretical physics, it is also applied in many areas such as modeling of biomolecules, material science and physical chemistry, etc. MD simulations are one of the tools used in the theoretical studies of biomolecular structures. Besides, MD simulations cannot achieve a result using only classical experiments; therefore, it is possible to obtain information about molecular structure and dynamics by using experiments such as NMR, X-ray crystallography and CD. [21] If atoms and molecules are allowed to interact for a certain period of time, the forces between the particles, intrinsic potential energy and molecular mechanical force fields are calculated from Newton's equations of motion and their trajectories are found. The time-dependent behaviour of the molecule gives information about the dynamics of molecular systems. Because molecular systems are complex systems consisting of a large number of particles, the analytical calculation is not possible without using computer-based methods. MD simulations can analytically determine the properties of systems with the help of numerical methods, and make them computable by creating parameters and algorithms for more complex molecules. Mathematically ill-conditioned MD simulations in relation to parameters can cause cumulative errors that cannot be completely eliminated but minimized. [22]

1.3.1. Classical Mechanics

MD simulation can be explained by Newton's second law or the motion equation from the basic laws of classical mechanics. Newton's second law according to the equation of motion;

$$F = m\,a$$

When the force (F) on each particle in the system is known, their accelerations (a) can be determined using their mass (m). When the motion equation is integrated, the time-varying positions, velocities and accelerations can be defined in the trajectory. According to the generated trajectory, the averages of these properties can be determined. In the meantime, it is possible to estimate the properties of the system from the past to the future at any time within the knowledge of the position and speed of each particle due to the fact that this method is deterministic. The force expressed depending on the potential energy (V):

$$F = - dV / dr$$

When the two equations are integrated, the equation of Newton's motion can be related to the change of the position in potential energy over time:

$$- dV / dr = m\, d^2r / d^2t$$

The acceleration can also be given as a derivative of the potential energy associated with the position:

$$a = -1/ m\, dV / dr$$

As a result, it is sufficient to know the acceleration that depends on the initial positions of the particles, the initial velocity distributions, the potential energy gradient to calculate a trajectory. Experimental analysis such as X-ray crystallography and NMR spectroscopy can be used to determine initial positions.

Usually, initial velocities are randomly selected from a Maxwell-Boltzmann or Gaussian distribution at a given temperature so that overall momentum does not occur. Maxwell-Boltzmann and Gaussian formula:

$$p\,(V_x) = (m / 2\pi\, k_B\, T)^{1/2} \exp\,(- ½\, m\, v_x^2 / k_B\, T)$$

The given formula indicates that any atom has a V_x velocity in the x direction at the T temperature as well as m is the mass, k_B is Boltzmann constant. [22]

1.3.2. Integration Algorithms

Potential energy is a function consisting of positions (3N) of all atoms in the system. Therefore, the equation of motion is not analytical solution but can be solved numerically. With several algorithms in computational methods, the acceleration depends on the gradient of the potential energy can be calculated with the help of force fields. Some of these algorithms are known as Integration Algorithms: Verlet, Leap-frog, Velocity Verlet and Beeman's algorithms. Whichever algorithm should be used depends on a few substances such as energy and momentum conservation, and computational efficiency. All integration algorithms use Taylor series expansion since it is assumed that the properties like position, speed and acceleration of the particle can be approximated.

$$r(t + dt) = r(t) + v(t)\, dt + ½\, a(t)\, dt^2 + \ldots$$

$$v(t + dt) = v(t) + a(t)\, dt + ½\, b(t)\, dt^2 + \ldots$$

$$a(t + dt) = a(t) + b(t)\, dt + \ldots$$

When using the Taylor series expansion, it represents that r is position, v is velocity in the first derivative with respect to time, and a is acceleration in the second derivative with respect to time. For example, to derive the Verlet algorithm:

$$r(t + dt) = r(t) + v(t)\, dt + ½\, a(t)\, dt^2$$

$$r(t – dt) = r(t) – v(t)\, dt + ½\, a(t)\, dt^2$$

can be written. When above two equations are summed:

$$r(t + dt) = 2r(t) – r\,(t – dt) + a(t)\, dt^2$$

According to the Verlet algorithm, the position and momentum at time t and the position at time t - dt should also be used to find the position at t + dt. Explicit speeds are not available in this algorithm. The Verlet algorithm has a medium sensitivity, although it is simple and does not require a large area for storage. Not every algorithm can be used for every analysis. They should be tested to make sure that the most appropriate algorithm is selected and the closest result to the truth is achieved. [22]

1.3.3. Empirical Force Field Models

Theoretical studies are needed to study the relationship between the structure, function and dynamics of biological molecules at atomic level. It is not possible to analyze problems in biological systems using quantum mechanics due to multi atom systems. However, much more reasonable results can be obtained when analyzing on empirical potential energy functions requiring less computation. (23)

The current generation force fields, also known as potential energy functions, provide a reasonable deal between accuracy and calculation efficiency. In addition, they are calibrated according to experimental results from X-ray crystallography and NMR and quantum mechanical calculations. The progression of the parameter sets of the potential energy functions is a laborious process that requires extensive optimization. The most commonly used force fields: AMBER (Assisted Model Building with Energy Refinement) [24], CHARMM (Chemistry at Harvard Macromolecular Mechanics) [25], GROMOS (Groningen Molecular Simulation) [26], and OPLS (Optimized Potentials for Liquid Simulations) are some of the examples given for potential energy functions. Therefore, force fields are empirical equations that are used in MD simulations for calculating potential energy, including the position of the particle of the molecule involved and its interaction with other particles.

1.3.3.1. Molecular Interactions

Energy, E, is a function usually expressed in Cartesian coordinates, including the atomic positions (R) of all atoms present in the system. As discussed in the previous section, the interactions used to calculate the potential energy can be examined in two categories: internal or bonded and external or non-bonded interactions. The sum of these two is equal to the total potential energy of the system.

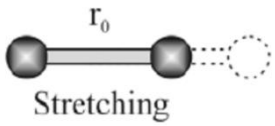
$$V = E_{bonded} + E_{non\text{-}bonded}$$

The energy of bonded interactions is expressed by the sum of the bond stretch, angle bend and torsion angle, which defines the bonds, angles and rotations of a molecule. [22]

$$E_{bonded} = E_{bond\text{-}stretch} + E_{angle\text{-}bend} + E_{rotate\text{-}bond\text{-}bond}$$

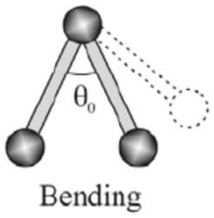
The first term of formula, $E_{bond\text{-}stretch}$, represents the interaction between two atoms connected by covalent bond. The calculated energy is a harmonic potential. Therefore, the ideal bond length used in the equation, b_0, is included in the displacement function and K_b is the force constant that is specific for each pair of bonded atoms.

$$E_{bond-stretch} = \sum_{1,2\,pairs} K_b \left(b - b_0\right)^2$$



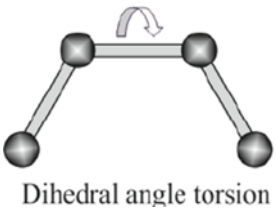
The second term of formula, $E_{angle\text{-}bend}$, is a harmonic potential energy. The deviation of the ideal angle value Q_0 is calculated from the change function. Q_0 and K_Q vary depending on the chemical type of the 3 atoms that make up the angle. [22]

$$E_{bond-bend} = \sum_{angles} K_\theta \left(\theta - \theta_0\right)^2$$



The third term calculates the potential energy of the system over atoms connected by 3 covalent bonds along a torsion at a dihedral angle. Torsion indicates the presence of angle steric barriers between atoms. Its function is periodic and is often expressed by the cosine function.

$$E_{rotate-along-bond} = \sum_{1,4\,pairs} K_\phi \left(1 - \cos\left(n\phi\right)\right)$$



On the other hands, non-bonded interactions are also observed between atoms without covalent bonds or between distant atoms with more than 3 covalent bonds. It is also expressed by the sum of Van der Waals interaction energy and electrostatic interaction energy. [22]

$$E_{\text{non-bonded}} = E_{\text{van-der-Waals}} + E_{\text{Electrostatic}}$$

Van der Waals interactions, which are very important for the stability of biological macromolecules, occur when the attractive and repulsive forces between the two atoms are in balance. The distance (r) decreases as the two atoms approach each other. As shown in Figure 2, the equilibrium distance at which the potential energy approaches the minimum equals the sum of the radii of two atoms. If the distance between atoms is less than the equilibrium distance, repulsive forces are generated due to the electron distributions and potential energy begins to increase. Likewise, if the distance between them increases, the attractive force electron cloud causes sudden dipole to induce dipole induction. as a result, potential energy will increase.

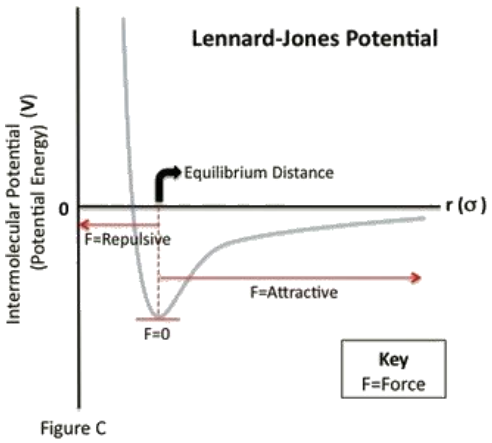


Figure 2: Potential Energy of van der Waals interaction can change according to distance between both atoms. Equilibrium distance equal to sum of radii of two atoms. [Reproduced without permission from libretext.org]

The Van-der-Waals interaction uses the Lennard-Jones potential, as follows:

$$E_{van-der-Waals} = \sum_{\substack{nonbonded \\ pairs}} \left(\frac{A_{ik}}{r_{ik}^{12}} - \frac{C_{ik}}{r_{ik}^{6}} \right)$$

Finally, the electrostatic interaction of a pair of atoms can be calculated using the function of the coulomb potential. D is the dielectric function for the environment, qi and qk are the charges of two atoms and r is the distance between them.

$$E_{electrostatic} = \sum_{\substack{nonbonded \\ pairs}} \frac{q_i q_k}{D r_{ik}}$$

Although the above mentioned molecular mechanics calculations are common for empirical force fields modeling, there are differences in the calculation of bonded and non-bonded

parameters. In addition, the potential energy functions do not include entropic contributions, in which case the minimum potential energy cannot fully meet the actual result. In experiments, equilibrium state, which is generally performed at constant temperature and pressure, can be associated with minimum Gibbs Free energy.

1.3.4. Role of Solvent in Molecular Dynamics Simulations

Solvents, water is generally used for this, have a huge impact on the structure, dynamics and thermodynamics of biomolecules, due to electrostatic interactions between the two charges. So, it can be showed with help of Coulomb's Law for the electrostatic interaction between two charges:

$$V_{elec} = \frac{q_i q_j}{\varepsilon_{eff} r_{ij}}$$

While q_i, q_j are the partial atomic charges, ε_{eff} refers to the effective dielectric constant as well as r_{ij} is the relative distance between the two charges. [22] The effects of the solvent can be incorporated at several different levels. One is simply adding a dielectric constant to the formula when calculating the potential energy of electrostatic interactions. In this case, the simulation of water molecules in the implicit treatment of the solvent is not included, but with an effective dielectric coefficient, $\varepsilon_{eff} = r_{ij}\,\varepsilon$, which is usually between 4 and 20, is calculated depending on the distance. This approach makes more sense than using unscreened partial charges. In another method, the explicit solvent model is prepared by assuming ε=1 by explicitly incorporating each solvent molecule. Clearly incorporated molecules provide electrostatic shielding. In this approach, it is necessary to prevent diffusion in order to keep the solvent molecules in the defined boundary conditions and to use a limited number of solvents to observe macroscopic properties. [27] Periodic boundary is a periodic system consisting of repetition of identical cell in all directions. When applying periodic boundary conditions, the molecule is covered by a central box one of the identical cells and surrounded by eight adjacent cells. In this case, the molecules of adjacent cells can interact with each other and can replace each other, while preserving the equivalent atom system. In another method, to reduce cost, the solvent effect is calculated by taking the perimeter of the protein known as solvation shells or by introducing a certain portion of proteins into a water sphere which is called active site solvation. [28]

1.4. α-Lactalbumin-derived peptide

The α-lactalbumin protein, also known as LALBA, is encoded by the human *LALBA* gene. [29] It is a protein that regulates lactose production in milk in most mammals. [30] α-lactalbumin consists of 123 amino acids in total. In residues 1-39 and 81-123, a primarily helical domain (α-domain) structure is observed, while a sheet / coil domain (β-domain) is observed in the rest of the molecule as in the Figure 3. [31]

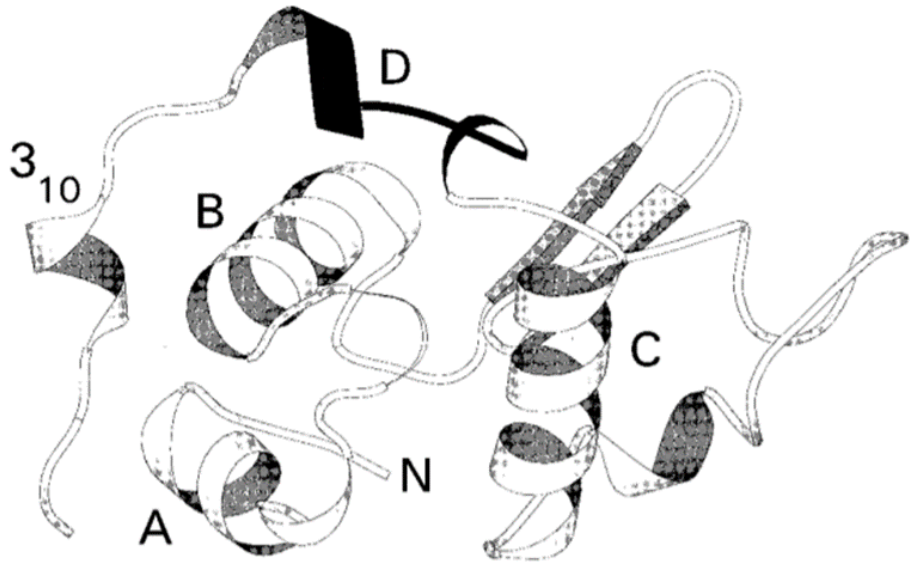


Figure 3: X-ray crystal structure of human α-LA. The α-domain structures are showed A, B, C and 3_{10} as white colour and 101-111 residues of α-LA is showed D as black color. [31]

In this thesis, the derived peptide form INYWLAHAKAG of the 101-111 residues of α-La protein determined by Mitsugu Araki and Atsuo Tamura will be used and this peptide will be called α-La peptide throughout the thesis. [32] The presence of stable 3_{10}-helical structure at the N-terminus (3-6 residues) and a generally disordered structure at the C-terminus of the peptide was determined by two independent NMR determinations. [33]

1.5. Purpose of the present thesis

The present thesis aims to validate the AMBER-15SB force field by molecular dynamics simulation of an α-Lactalbumin-derived peptide. By trajectory analysis of AMBER-15SB with other force fields which have previously been subjected to molecular dynamics simulation, to compare whether this is the closest to the experimental results.

2. METHODS

2.1. Peptide and Design of Simulations

The INYWLAHAKAG peptide defined by Araki and Tamura, one of the two well-defined sequence variants for the α-La peptide, was selected for simulation. Since it contains additional experimental results of circular dichroism spectroscopy, except for the NMR result, and the structure of peptide in the solution is more suitable for interpretation. [32]

Two force fields AMBER99SB-STAR-ILDN and AMBER99SB-4fs-STAR-ILDN determined based on previous studies will be used for validation of AMBER15SB. They will then be referred to as STAR, 4fs-STAR, and 15SB respectively. The STAR force field is predicted to obtain a more accurate result by obtaining the frequency of motion of the Pro ring interconversion 6 times faster than the AMBER99SB force field due to improved Pro and Hyp parameters derived from NMR J-coupling and experimental correlation times. [34] 4fs force fields are examples of longer-time-step MD trajectories. Hydrogen mass repartitioning (HMR) method is used to accelerate molecular dynamic simulation. It is aimed to minimize the high-frequency movements of molecules by repartitioning the masses of heavy atoms to the hydrogen atoms to which they are attached. The mass distribution of the system does not affect the result, but two limitations should be observed: simulation stability and formal discretization error. In summary, longer-time-step MD trajectories are used to increase the time step of simulations for more efficient results. [35]

2.2. System preparation and simulation protocol

A cluster of computers connected in parallel provides a high computational power requirement for MD simulations. Norma, an example of a computational cluster, is used in this experiment. The NAMD program, which is a software used for the simulation of biomolecular systems, was used for the simulation of α-La molecular dynamics. The NAMD software is compatible with the force fields used. [36] Three files required for simulation with NAMD: a .pdb file from the protein data bank containing the coordinates and velocities of the atoms in the system, a customized file for AMBER force fields containing all the necessary parameters (in PRMTOP format for AMBER), and a configuration file containing all the necessary information for simulation.

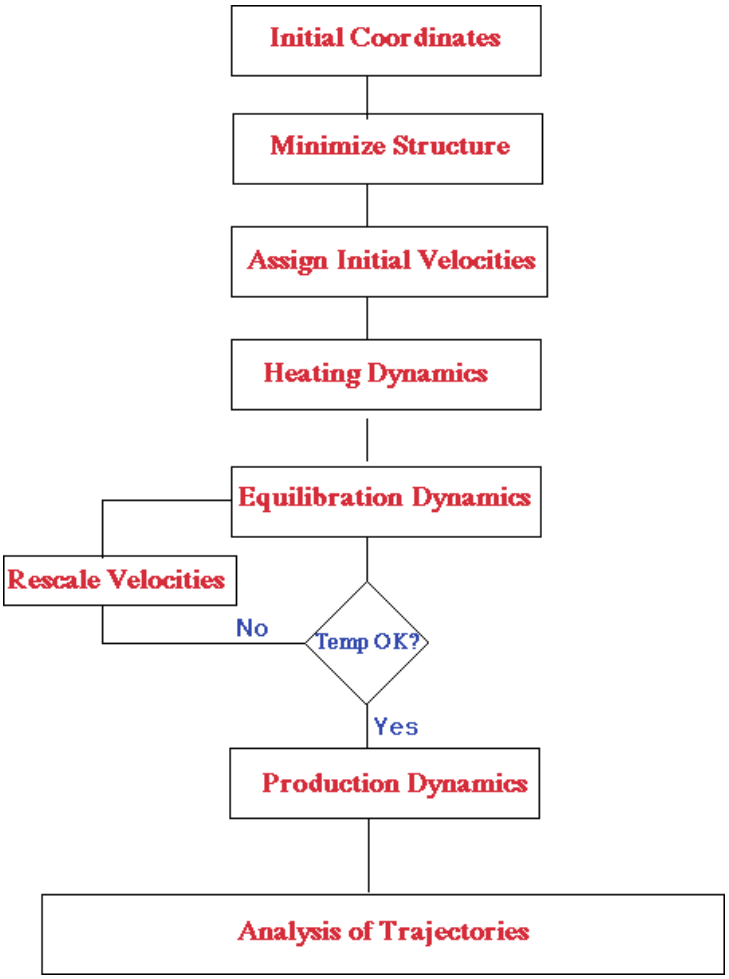


Figure 4: Steps of Molecular Dynamic Simulation. (was taken without permission from tutorial of CHARMM MD simulation tutorial, http://www.ch.embnet.org/MD_tutorial/pages/MD.Part3.html)

The steps in MD simulation are respectively as follows: Initial coordinates generally obtained based on X-ray diffraction or NMR analysis, energy minimization by removing strong Van der Waals interactions, assigning initial velocities with the lowest temperature available, heating dynamics until the ideal temperature reaches, the equilibrium phase associated with Newton's second subject when the structure, pressure, temperature and energy of the peptide are fixed, if the temperature is not suitable rescale velocities, productive phase which is the input generation of the coordinates, velocities and energy of the atoms recorded in equilibrium phase, and analysis of trajectories is the final step for evaluating the results as in Figure 4. [22]

The molecular dynamic simulations mentioned in this thesis have already been performed. The necessary information about the simulations are as follows. First minimization was performed with 1000 conjugate gradient steps. The number of time and frames of the three folding simulations performed by the NAMD program are respectively; 6,2 µs and 6271300 for 15SB, 2,5 µs and 2500837 for STAR, and 3,5 µs and 3501200 for 4fs-STAR. The Langevin dynamics

applied by the NAMD program has kept the temperature and pressure under control. Final temperature is 320 K and pressure is 1 atm.

2.3. Trajectory analysis

The data obtained from the MD simulation was carried out in the *carma* [37] and *grcarma* [38] programs using the graphical interface for the analysis. .psf and .dcd files, which are two output files for each of the trajectories after simulation, are used as input in these programs. The PSF file (protein structure file) contains all the topological information of the system, namely atoms, bonds, angles, and also the mass and charge of each nucleus. [39] The DCD file is a binary file that contains the coordinate sets of all trajectory atoms associated with a frame at a time. [40] Several analyzes were carried out separately for three trajectories with the CARMA program: calculation of frame-to-frame RMSD matrices, calculation of secondary structure assignments with STRIDE, dihedral space principal component analysis.

In addition, two calculations were made for the comparison of Nuclear Overhauser Effect (NOE), one of the NMR observables. First of all, the distance between atoms was calculated by using *carma* program by using .psf and .dcd files. Then program c in **Appendix (A1)** was run and NOE values were calculated. The second C program in **Appendix (A2)** was run to determine the upper bound violation values using the distance values (r) -6 from the result obtained with the experimental NOE values. The total number of violations between interacting atoms and the average upper bound violation were calculated by dividing all observed upper bound violation values by total proton pairs. Based on the analysis and calculations obtained, it was interpreted for 15SB validation.

3. RESULTS

3.1. Introduction

After MD simulations were performed using STAR and 4fs-STAR force fields to compare the validation of 15SB force field on α-La peptide, analyzes were performed with *carma* and *grcarma* simulation programs. The .psf and .dcd files obtained from the simulations were used for analysis programs. The results of the analyzes are given in the following sections.

3.2. RMSD Analysis

The root-mean-square deviation (RMSD) is a Molecular Dynamics simulation data analysis commonly used in Structural Biology and Bioinformatics, which calculates the average distance between backbone atoms of superimposed proteins using atomic positions. [42] RMSD calculation formula is as follows:

$$\mathbf{RMSD} = \sqrt{\Sigma(x_i - x_{ref})^2 / N}$$

X_i in the formula gives the coordinates of the atoms at the specific time, while X_{ref} gives the atomic coordinates of the reference structure as well as N is the number of atoms. [43] These two structures must be superimposed. The smaller the calculated RMSD, the greater the similarity between the structure taken from the specific time and the referenced structure. If the RMSD result is 0.0 Å, the two structures are said to be identical. The results of the RMSD are indicated by the colors expressed as magnitude. The blue color in the RMSD matrix means low RMSD's magnitude, yellow color means medium RMSD's magnitude, and red color means high RMSD's magnitude. [44] The blue line formed on the diagonal line of the RMSD matrix indicates a stable structure during that time. RMSD matrixes belonging to three different force fields produced by GRCARMA are shown in Figure 5.

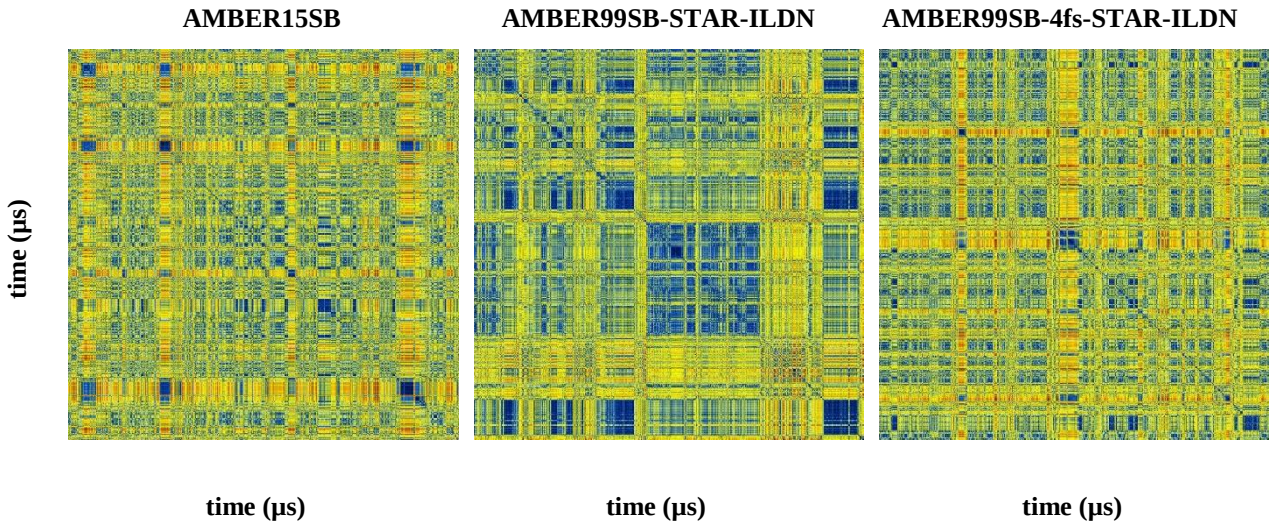


Figure 5: RMSD matrixes of Cα atoms between the 101-111 residues of the α-lactalbumin peptide, performed from left to right with the AMBER15SB, AMBER99SB-STAR-ILDN, and AMBER99SB-4fs-STAR-ILDN force fields, respectively.

The RMSD matrix of one of the trajectories belongs to 15SB force field, shows a total of 6.2 μs with 6271300 frames, while the matrix of STAR force fields' trajectory shows 2.5 μs with 2500837 frames, and the matrix of 4fs-STAR force fields' trajectory shows 3.5 μs in total with 3501200 frames. The force fields were compared using the 101-111 residues of α-La, while the specific stable structure was observed with the STAR force field during the simulation time. In the RMSD matrix of the STAR force field, more blue color, ie low RMSD magnitude, was measured. In the RMSD matrix of the force field of 15SB, yellow colors are more intense, ie the medium RMSD magnitude, is measured. In the RMSD matrix of the 4fs-STAR force field, red color was observed in some regions while blue and yellow color areas are intensity. In this case, the most stable structure of α-La was observed by STAR force field.

3.3. Secondary Structure Analysis

This analysis contributes to the understanding of the tertiary structures that may occur and the examination of the folding process by calculating all secondary structures that may occur on the peptide during the simulation. The STRIDE (Structural Identification) method, which produces an algorithm using the dihedral angle and hydrogen bond energies of the backbone of the peptide, is used for this study. [45] As a result of the analysis performed in GRCARMA program, the secondary structures of the peptide are colored by a text file produced by STRIDE algorithm. The graphics prepared using the trajectories of the three simulations are shown below.

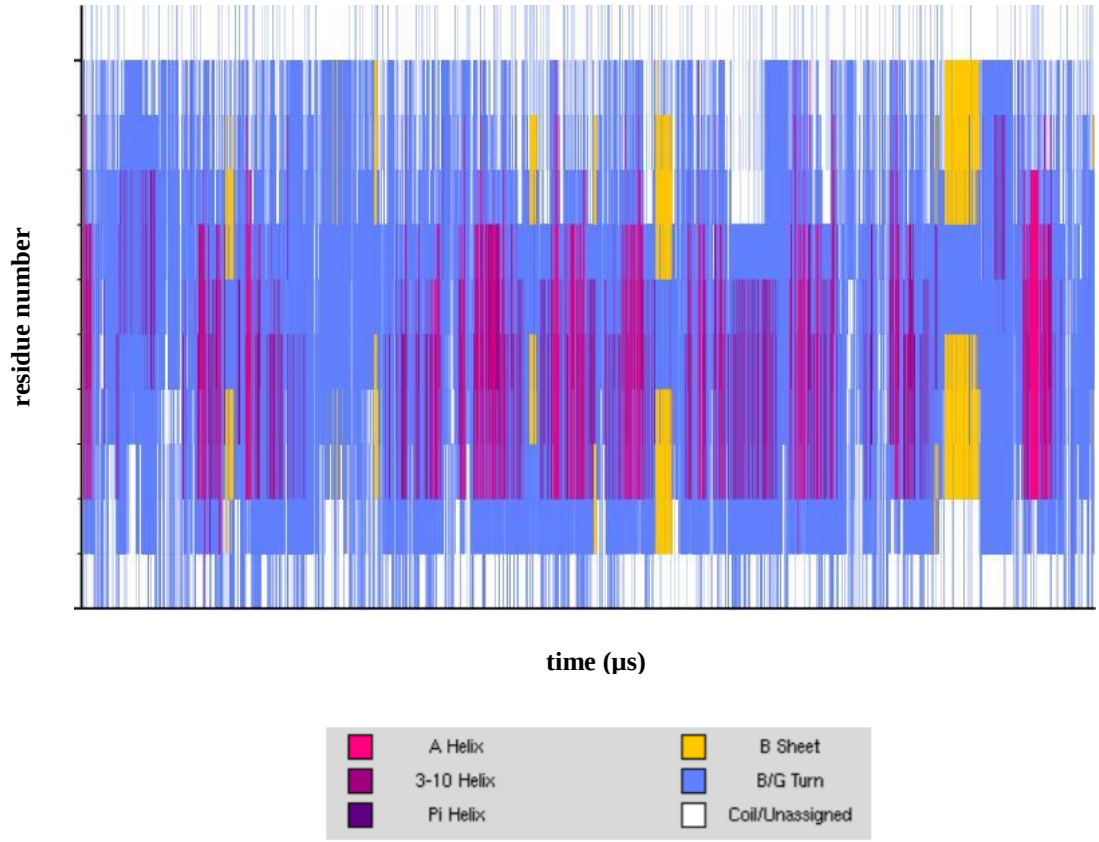


Figure 6: Secondary structure prediction of the trajectory simulated using the AMBER15SB force field. In the analysis, secondary structures are expressed as colored: pink color for α-helix, purple color for 3-10 helix, dark purple color for pi helix, yellow color for β-sheet, blue color for β-turn, and white color for random coil.

Figure 6 shows the secondary structure prediction analysis of the simulation using the 15SB force field. The vertical axis of the graph indicates residues from 101 up to 111 from bottom to top, while the horizontal axis represents the simulation time from 0 to 6.2 µs from left to right. As a result of the analysis, β-turn structure is mostly observed but this structure is not stable. Sometimes, instead of this structure, the formation of β-sheet was observed. In addition, 3_{10}

helix structures are frequently observed between 103-108 residues -YWLAHA-, while a helix structure is rarely observed in the same residues.

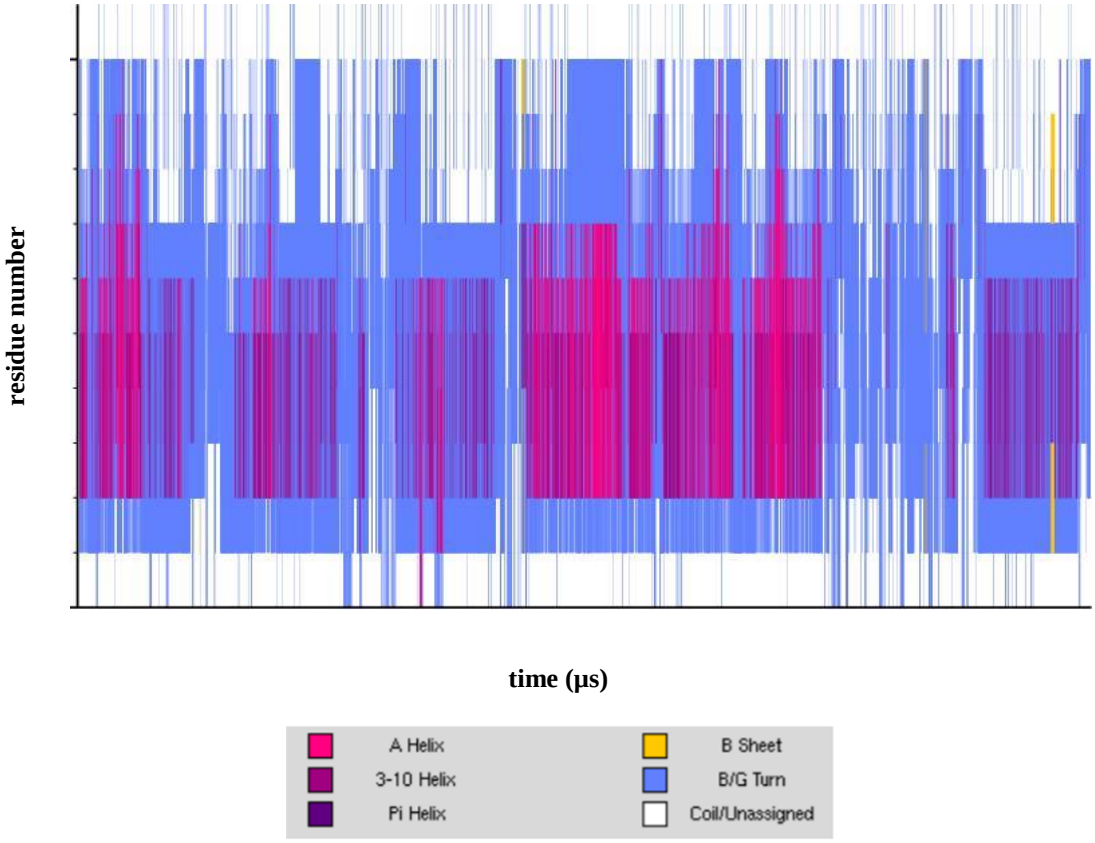


Figure 7: Secondary structure prediction of the trajectory simulated using the AMBER99SB-STAR-ILDN force field. In the analysis, secondary structures are expressed as colored: pink color for α-helix, purple color for 3-10 helix, dark purple color for pi helix, yellow color for β-sheet, blue color for β-turn, and white color for random coil.

Figure 7 shows the secondary structure prediction analysis of the simulation using the STAR force field. The vertical axis of the graph indicates residues from 101 up to 111th from bottom to top, while the horizontal axis represents the simulation time from 0 to 2.5 µs from left to right. At the end of the analysis, it was observed that the peptide was more stable than other force fields. While mostly β-turn and random coil structures were observed between the 101-111 residues of the peptide, helix structures between the 103-108 residues -YWLAHA- were observed.

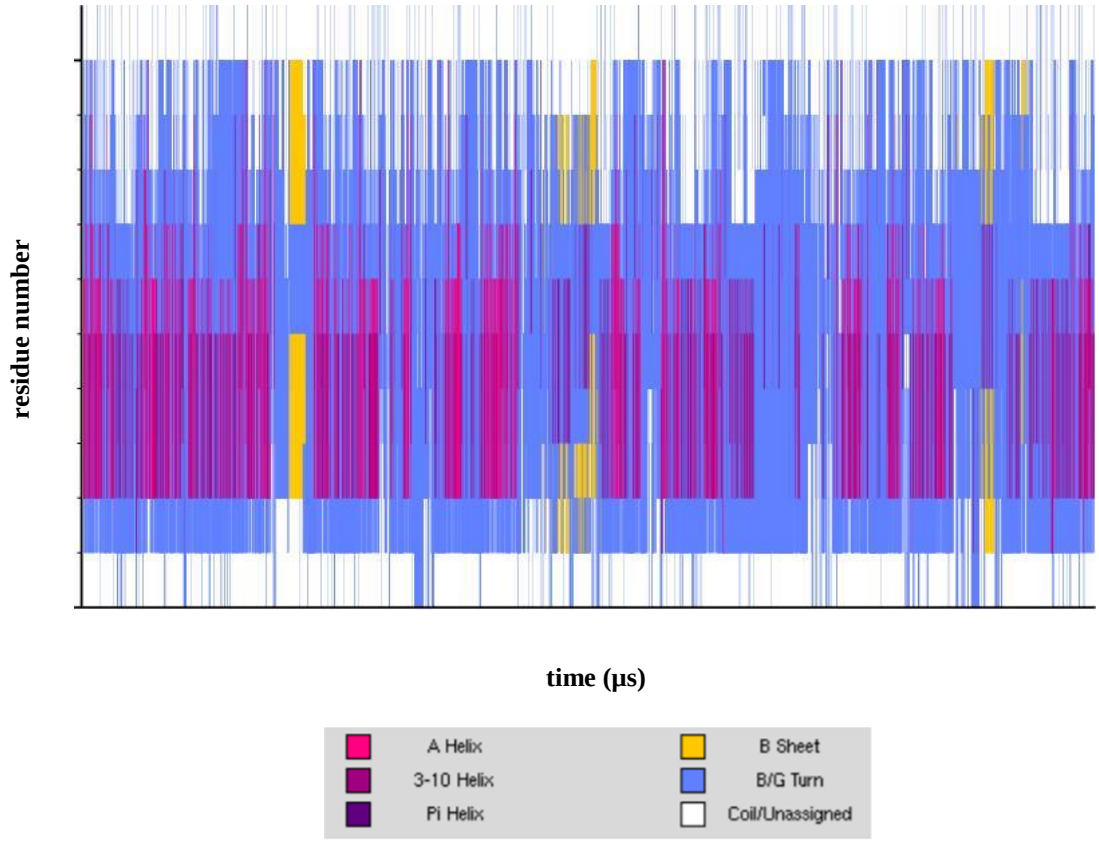


Figure 8: Secondary structure prediction of the trajectory simulated using the AMBER99SB-4fs-STAR-ILDN force field. In the analysis, secondary structures are expressed as colored: pink color for α-helix, purple color for 3-10 helix, dark purple color for pi helix, yellow color for β-sheet, blue color for β-turn, and white color for random coil.

Figure 8 shows the secondary structure prediction analysis of the simulation using the 15SB force field. The vertical axis of the graph indicates residues from 101 up to 111th from bottom to top, while the horizontal axis represents the simulation time from 0 to 6.2 µs from left to right. As a result of the analysis, during the simulation period, β-turn and random coils were observed non-continuously throughout the peptide, while α-helix and 3-10 helix structures were frequently observed between 103-107 residues -YWLAH-. In addition, between 102-106 residues -NYWLA- and 108-111 residues -AKAG-, although rarely, β-sheet structure was observed.

For easier comparison, the RMSD matrices and secondary structure prediction prepared using STRIDE algorithm are combined and shown again in Figure 9.

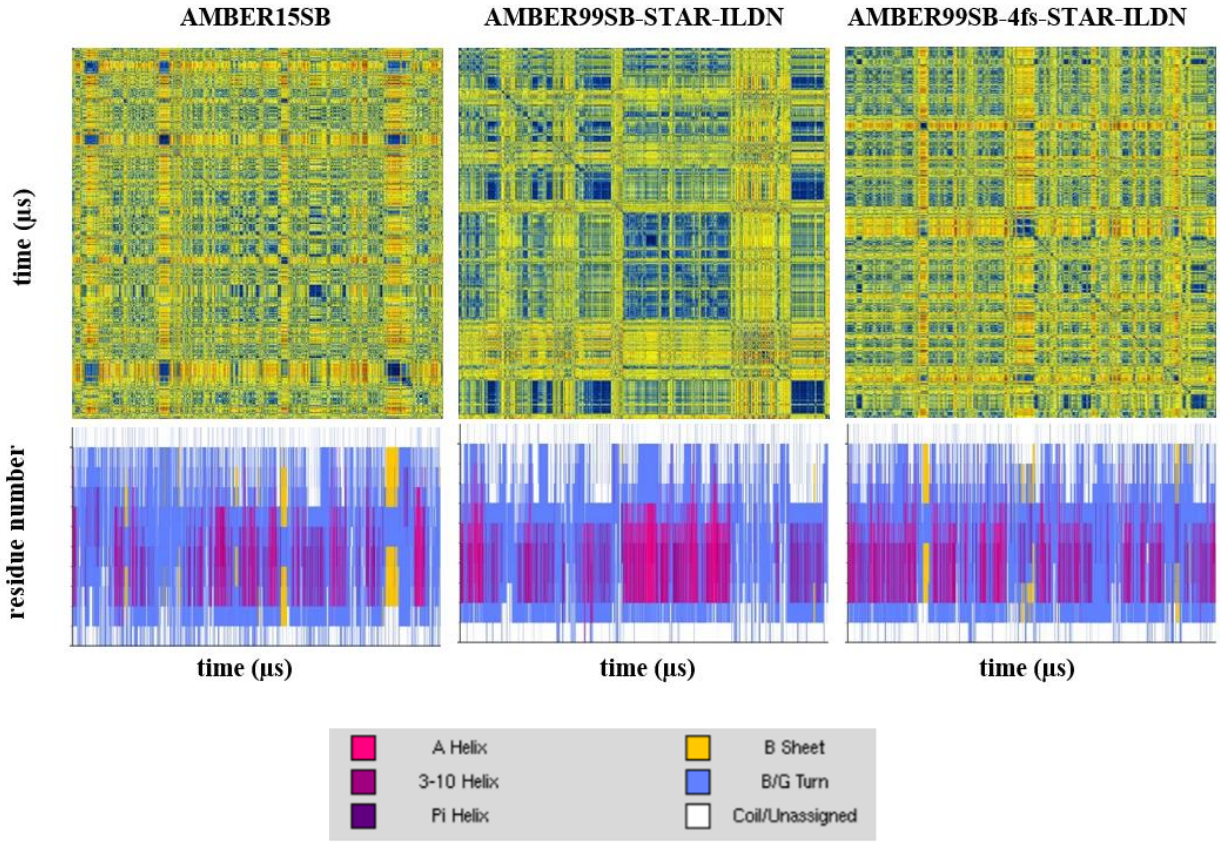


Figure 9: RMSD and secondary structure analysis, which is a trajectory analysis of simulations performed with different force fields.

Firstly, the horizontal axes indicate the time for both analyzes, so the two analysis types of each force field are available to be interpreted together. For example, it is possible to comment on the stability of the peptide in which the secondary structure is formed. The analysis results of the 15SB force field on the left side of Figure 9 show that when the β-sheet, β-turn and random coil are observed, the RMSD magnitude value increases that means the stability of residues 101-111 of the α-La peptide is reduced. In addition, during the simulation period when α-helix and 3-10 helix are observed, the blue color is observed in the RMSD matrix. This means that the peptide structure remains stable for a short time. When the analysis results of the STAR force field shown in the middle of Figure 9 are analyzed, the blue color is seen in the RMSD matrix which is equivalent to the simulation times showing the formation of α-helix, 3-10 helix, and pi-helix from secondary structure analysis. According to other results, the most stable structures are these. In addition, yellow color is observed when β-turn and random coil formation occurs. In others, the formation of these structures showed a high RMSD magnitude.

Finally, when the analysis results of the 4fs-STAR force field are examined, a blue color indicating the stability of the RMSD matrix is observed in the same time as α-helix and 3-10 helix formations. During simulation times where the β-sheet structure is observed, the RMSD magnitude value increases and is expressed in red color. In summary, RMSD and secondary structure analysis showed that different helix formations were more stable than β-sheet formations of 101-111 residues of α-La peptide. 3-10 helix and α-helix secondary formations are observed to be not pure. The residues (i) and (i + 3) and (i+ 4) residues between the bifurcated hydrogen bonds formed precisely neither α-helix nor 310 helix structure was observed to be stable.

3.3. Dihedral Angle Principal Component Analysis (dPCA)

Principal component analysis (PCA) is a transformation technique that reduces the dimensionality of a high dataset containing a large number of related variables to a smaller size while preserving the data in the dataset. In other words, PCA is a mathematical technique of explaining information in a multivariate dataset with fewer variables and minimum information loss. The variables obtained after the transformation are called the basic components of the first variables. In this way, PCA is widely used in the field of Molecular Dynamics, especially because it is able to identify various datasets and distinguish their similarities and differences.

Dihedral angle Principal Component Analysis (dPCA) provides a systematic approach to the structure by creating a low-dimensional free energy landscape. Unlike cartesian Principal Component Analysis (cPCA), dPCA provides a more efficient representation of the conformational dynamics of small peptides. The basic understanding of dPCA in molecular dynamics is the representation of correlated internal movements with covariance matrices. [50] The dPCA shows the free energy landscape in more detail by separating the substantial internal motion from the insignificant overall motion. The dPCA method uses dihedral angles (φi, ψi) of the backbone of the peptide, as opposed to cPCA. The reason for this is that the observation of the change of the large amplitudes as opposed to the internal coordinates such as bond angles and bond lengths can be considered as looking at the big picture. Since the dihedral angles are mentioned in dPCA, it is circular and periodic since it can rotate 360 degrees. In the cPCA, since regular data is used to define a metric, cos- and sin- transformed must be used to form recovered a metric coordinate when defining distances between atoms. After the transformation, the mean and the covariance matrix are calculated by forming the diagonal matrix to create eigenvectors (v^i) and eigenvalues (λ_i). Eigenvectors with the largest eigenvalues (λ_1) are considered as major components and others are ordered by size. Principal component

is calculated using the data $\mathbf{q} = (q_1, \ldots, q_{3M})^T$, reflect the free energy surface of the system, with the following equation.

$$V_i = \mathbf{v}^{(i)} \cdot \mathbf{q}$$

Biological processes, especially protein dynamic studies, are defined by free energy as in the following equation. Restricted to two dimensions:

$$\Delta G(V_1, V_2) = -k_B T[\ln \rho(V_1, V_2) - \ln \rho_{\max}]$$

ρ is estimation of possible density function obtained from the data, $\rho_{\max}$ indicates the maximum of this value. $\Delta G = 0$ denotes the lowest level of free energy and the aim is to approach this value, so subtraction is performed as in the above equation. In this way, the data obtained from free energy landscape shed light on protein folding studies.[41]

With the GRCARMA trajectory analysis program, free energy landscapes of trajectories belonging to the three different force fields were observed by dPCA method. The two-dimensional free energy landscapes are shown in Figure 10.

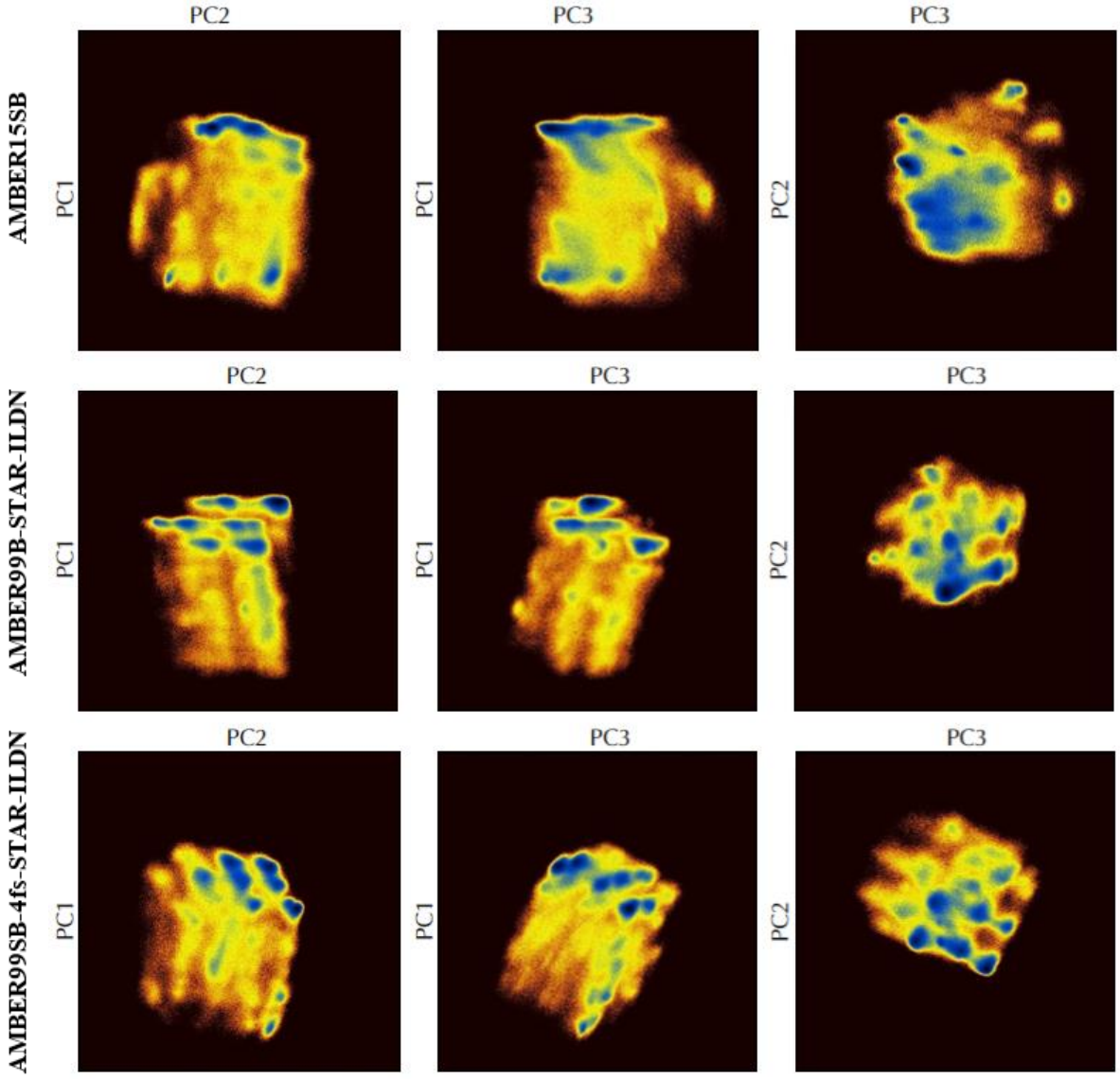


Figure 13: The free energy landscapes of trajectories belonging to three different force fields of the 101-111 residues of the α-lactalbumin peptide obtained by the dPCA method are presented in two dimensions. The first three principal components along the line belong to the same force field. ΔG plots of trajectories of AMBER15SB force field in the first column, AMBER-99SB-STAR-ILDN force field in the second column, and AMBER99SB-4s-STAR-ILDN force field in the third column. The blue regions appearing in the diagram refer to various energy minima.

The free energy landscapes belonging to residues 101-111 of α-La peptide were observed for three different force fields. In the diagram generated by dPCA analysis, the dark blue regions represent low free energy in the landscape like global minima and this scale increases the free energy from blue to yellow and yellow to red region. Each landscape is shown from three different angles as shown each line. Looking at the result, the first observation is that the landscape structure is similar in all three trajectories. In the diagrams, blue regions were observed scattered due to the uncertainty of the energy minima bounder. However, this region

would have to be more sharp and distinct to obtain a stable state of the peptide. In this case, many unstable intermediate state occurrences, similar to the stable structure, can be observed during the simulation.

3.5. NMR observables and their comparison with simulation

There is a mutual relationship between molecular dynamics simulations and experiments. On the one hand, experimental fields are used while force fields and simulations are analyzed. On the other hand, simulations are used to interpret experimental results through atomic visualization. Nuclear Magnetic Resonance (NMR) is known as an experimental method used to describe the structure and dynamics of biomolecules and also is a physical phenomenon dependent on the magnetic properties of atomic nuclei. The odd-numbered nucleon, usually 1H is used, and some even-numbered nuclei have a magnetic moment. NMR affects the magnetic moment of the nucleus with another magnetic field added from the outside and disrupts the effect of the electromagnetic wave, allowing the momentum to change direction for a moment. A resonance is observed at that frequency used in nuclear magnetic resonance spectroscopy to obtain physical, chemical and structural information about the molecule. [46] NMR observables are generally calculated using trajectories of simulations and they are compared with experimental data. Some of the many phenomena originating from NMR are: nuclear Overhauser effect, chemical shifts, and J-coupling. In this thesis, the calculation of NOE analysis from NMR observables will be shown in this section and the comparison with previous results will be shown in the other section. [48]

3.5.1. Nuclear Overhauser Effect (NOE)

The Nuclear Overhauser effect (NOE) is defined as the observed change in the absorption intensity of nuclei A by changing the normal spin population due to radiation from another nuclei B close to it. NOE can be measured by changes in spin population of nuclei B. The Nuclear Overhauser effect is not directly related to the chemical bonds of the structure studied, but to their interatomic distances. The NOE signal can be received for atoms that are spatially close to each other. The observed NOE signals provide information about the three-dimensional structure of the molecule studied. When the inter-atomic distances are known due to NOEs, the bonds formed between the atoms associated with each other are more easily identified and can be separated from one. If the distance between the two atoms is greater than 5.5 Angstrom, the NOE signal cannot be received. [49] The following formula is used when calculating the NOE signal:

$$NOE = 1/r^6 f(t_c)$$

While r represents the distance between the two nuclei, while t_c represents the time required for 1 rad. The distances between atoms identified by NOEs signals give information about their distance restraints and can be interpreted using secondary structures. As shown in the equation, NOE is related to the -6 force of the inter-nuclear distance $(r)^{-6}$. For most large peptides, $(r)^{-3}$ is preferred in the equation, $(r)^{-6}$ is preferred for small peptides. NOE signals are classified according to the distance between atoms when interacting. NOE signals were categorized as Strong for 1.8-2.7 Å, medium for 2.7 - 3.3 Å, weak for 3.3-5.0 Å. [48] These are referred to as experimental classification S, M and W, respectively.

The atom pairs that could give NOE signal obtained from the simulation were compared with the atom pairs giving NOE signal obtained from experimental data before calculation. The 101-111 residues of the α-La peptide comprise a total of 178 atoms, including NOE signal among 138 proton pairs. After calculating the distances between atoms by CARMA program, $(r)^{-3}$ and $(r)^{-6}$ were calculated by running the C program script in **Appendix (A1)** based on the formula below:

$$R = (< R_{ij}^{-6} >)^{-1/6} \quad R = (< R_{ij}^{-3} >)^{-1/3}$$

The upper bound violation method is used to validate the NOE values obtained from the simulation with experimental data. This method helps to detect restraints caused by violating NOE values of the structure that is, it aims to validate the obtained data by finding inconsistencies. For a restraint to be mentioned, the range of $(r)^{-6}$ value must be greater than the upper bound of the given NOE value. [48] For example, the value of $(r)^{-6}$ of a proton pair was calculated as 3.4 Å and the NOE signal shows medium. If the NOE signal is medium, the upper bound value of a NOE value is 3.3 Å, while the lower bound value is 2.7 Å. In this case, the measured value $(r)^{-6}$ is above the upper bound value and the upper bound violation is determined. This value must be calculated for the average upper bound violation value. The violation is calculated by the following equation.

$$v(i,j) = r^{-6} - nmr(i,j)$$

The value of V (i, j) is the violation between the pairs expressed as i and j, and nmr(i, j) is the upper bound value from the experimental data. When the validation value of the sample given in the previous paragraph is calculated according to this formula, it will be 0.1 Å.

NOE calculations were calculated separately for simulation trajectories from three different force fields. The results were combined to compare with the $(r)^{-6}$ values of experimental data. Table 1 shown below shows pair, the $(r)^{-6}$ values of 15SB, STAR, and 4fs-STAR force fields, respectively, their upper bound violation values, proton numbers, experimental classification and residue numbers of two nuclei close to each other.

Table 1: The $(r)^{-6}$ values used in NOE analysis to validate simulations prepared with three different force fields, AMBER15SB, AMBER99SB-STAR-ILDN and AMBER99SB-4fs-STAR-ILDN with experimental data. Yellow color indicates NOE values with upper bound violation.

Pair	**15SB** $(r)^{-6}$	**Upper bound violation**	**STAR** $(r)^{-6}$	**Upper bound violation**	**4fs-STAR** $(r)^{-6}$	**Upper bound violation**	**Proton number**	**Residue number & exp. classification**
1	2.481071 S		2,462541 S		2,461213 S		3 - 6	1I - 1I W
2	2.583273 S		2,457837 S		2,461241 S		6 - 8	1I - 1I M
3	2.956636 M		2,917884 M		2,874216 M		6 - 11	1I - 1I W
4	3.079931 M		3,330569 W		3,362083 W		6 - 18	1I - 1I W
5	2.546668 S		2,548021 S		2,545642 S		8 - 11	1I - 1I W
6	2.685134 S		2,789152 M		2,78731 M		8 - 15	1I - 1I M
7	3.268227 M		2,835574 M		2,798298 M		8 - 23	1I - 2N W
8	3.149422 M		3,176046 M		3,184574 M		11 - 18	1I - 1I W
9	3.537251 W		4,005717 W		3,969294 W		11 - 23	1I - 2N W
10	4.523066 W		6,068139 W		5,91945 W		11 - 45	1I - 3Y W
11	4.699332 W		5,169561 W		5,255258 W		11 - 47	1I - 3Y W
12	1.734409 S		1,751424 S		1,749962 S		14 - 15	1I - 1I S
13	2.538267 S		2,548755 S		2,545186 S		14 - 18	1I - 1I W
14	5.287613 W		5,174417 W		5,167828 W		18 - 47	1I - 3Y W
15	7.394228 W	**3,314228**	6,070646 W	**1,990646**	6,407717 W	**2,327717**	18 - 60	1I - 4W W
16	2.839251 M		2,838034 M		2,872849 M		23 - 25	2N - 2N W
17	2.560532 S		2,528912 S		2,517562 S		23 - 27	2N - 2N W
18	2.802993 M		2,626269 S		2,655377 S		23 - 28	2N - 2N W
19	3.203585 M		3,349983 W		3,558538 W		23 - 37	2N - 3Y W
20	2.612762 S		2,575236 S		2,590415 S		25 - 27	2N - 2N S
21	2.595047 S		2,691438 S		2,675681 S		25 - 28	2N - 2N M
22	4.178218 W	**0,80218**	4,335325 W	**0,685325**	4,279774 W	**0,299774**	25 - 58	2N - 4W W
23	1.740596 S		1,752025 S		1,751077 S		27 - 28	2N - 2N S
24	3.420220 W		2,464163 S		2,47061 S		27 - 33	2N - 2N W
25	3.089009 M		3,641782 W		3,60783 W		27 - 37	2N - 3Y W
26	4.167309 W		3,986824 W		3,93695 W		27 - 86	2N - 5L W
27	2.563088 S		3,756933 W		3,745095 M		28 - 32	2N - 2N W
28	3.542161 W		2,675382 S		2,657931 S		28 - 33	2N - 2N W
29	3.186904 M		3,11418 M		3,072933 M		28 - 37	2N - 3Y W
30	1.735827 S		1,749397 S		1,749062 S		32 - 33	2N - 2N S
31	5.619146 W	**0,119146**	4,128271 W		4,127332 W		32 - 62	2N - 4W W

32	2.818651 M		2,811838 M		2,807048 M		37 - 39	3Y - 3Y M
33	2.903367 M		2,582184 S		2,606757 S		37 - 41	3Y - 3Y W
34	2.903319 M		3,027708 M		3,04339 M		37 - 45	3Y - 3Y W
35	2.607674 S		2,751451 M	**0,071451**	2,644531 S		37 - 58	3Y - 4W S
36	2.480405 S		2,484259 S		2,483744 S		39 - 41	3Y - 3Y W
37	3.341477 W		3,115395 M		3,117915 M		39 - 45	3Y - 3Y W
38	5.659331 W		5,424614 W		5,442091 W		39 - 47	3Y - 3Y W
39	3.755292 W		3,588957 W		3,641154 W		39 - 101	3Y - 6A W
40	3.418155 W		3,165843 M		3,233196 M		39 - 106	3Y - 6A W
41	2.766231 M		2,75336 M		2,762494 S		41 - 45	3Y - 3Y W
42	5.000275 W		5,01851 W		5,022608 W		41 - 47	3Y - 3Y W
43	3.097927 M		3,53509 M		3,50161 W		41 - 58	3Y - 4W W
44	4.812158 W		5,042621 W		5,034573 W		41 - 60	3Y - 4W W
45	2.454283 S		2,472979 S		2,471438 S		45 - 47	3Y - 3Y W
46	3.678810 W		3,778117 W		3,841277 W		45 - 58	3Y - 4W W
47	4.006906 W		4,106526 W		4,17221 W		45 - 60	3Y - 4W W
48	4.954850 W		5,103781 W		5,232162 W		45 - 62	3Y - 4W W
49	4.240334 W		4,346666 W		4,262186 W		45 - 77	3Y - 4W W
50	6.236425 W		5,738873 W		5,721197 W		45 - 73	3Y - 4W W
51	3.813563 W		3,938153 W		3,83725 W		47 - 75	3Y - 4W W
52	4.307263 W		5,03412 W		4,480706 W		47 - 86	3Y - 5L W
53	6.043411 W		5,951668 W		6,209247 W		47 - 106	3Y - 6A W
54	5.646613 W		5,134753 W		5,117577 W		47 - 121	3Y - 7H W
55	2.868849 M		2,881896 M		2,888709 M		58 - 60	4W - 4W M
56	2.908583 M		2,693594 S		2,744525 M		58 - 62	4W - 4W M
57	2.654615 S		2,694798 S		2,730372 M		58 - 63	4W - 4W M
58	2.703130 M		2,884226 M		2,804333 M		58 - 77	4W - 4W W
59	2.480391 S		2,587881 S		2,535512 S		58 - 82	4W - 5L M
60	2.469619 S		2,530664 S		2,520886 S		60 - 62	4W - 4W M
61	2.507130 S		2,504976 S		2,474096 S		60 - 63	4W - 4W M
62	3.508286 W		3,251679 M		3,330053 W		60 - 66	4W - 4W W
63	3.323182 W		2,969154 M		3,014963 M		60 - 77	4W - 4W W
64	4.308291 W	**0,198291**	4,165286 W	**0,055286**	4,19216 W	**0,08216**	60 - 101	4W - 6A W
65	4.148744 W	**0,878744**	3,934016 W	**0,864016**	3,886072 W	**0,616072**	60 - 115	4W - 7H M
66	1.727354 S		1,743049 S		1,741325 S		62 - 63	4W - 4W S
67	2.806634 M		2,784751 M		2,799047 M		62 - 66	4W - 4W W
68	3.155252 M		3,130289 M		3,077625 M		62 - 77	4W - 4W W
69	3.413223 W		3,412261 W		3,389986 W		62 - 82	4W - 5L W
70	3.068120 M		3,03979 M		3,030337 M		63 - 66	4W - 4W W
71	3.312272 W		3,246895 M		3,424608 W		63 - 82	4W - 5L W
72	2.559937 S		2,547845 S		2,547413 S		66 - 68	4W - 4W W
73	4.901184 W		5,291152 W		5,035814 W		66 - 92	4W - 5L W
74	2.452209 S		2,469274 S		2,469139 S		77 - 75	4W - 4W M
75	2.886166 M		2,863754 M		2,863458 M		68 - 71	4W - 4W W
76	2.467654 S		2,484915 S		2,484299 S		71 - 73	4W - 4W S
77	2.811329 M		2,861316 M		2,864107 M		82 - 84	5L - 5L M

78	2.535769 S		2,473858 S		2,484257 S		82 - 86	5L - 5L W
79	4.386062 W		4,184616 W		4,234179 W		82 - 92	5L - 5L W
80	2.522661 S		2,368415 S		2,385934 S		82 - 101	5L - 6A M
81	3.921565 W		4,091265 W		4,18119 W		82 - 111	5L - 7H W
82	2.651642 S		2,793281 M		2,795216 M		84 - 86	5L - 5L W
83	2.982991 M		2,984914 M		2,993937 M		84 - 89	5L - 5L W
84	2.949377 M		3,202901 M		3,221246 M		84 - 92	5L - 5L W
85	2.992649 M		2,779591 M		2,77437 M		84 - 96	5L - 5L W
86	5.215048 W		5,554625 W		5,446133 W		84 - 106	5L - 6A W
87	4.114622 W	**0,414622**	3,992602 W	**0,292602**	4,033632 W	**0,333632**	84 - 111	5L - 7H W
88	2.809730 M		2,808839 M		2,798771 S		86 - 92	5L - 5L W
89	3.104759 M		3,29539 M		3,312596 W		86 - 96	5L - 5L W
90	3.290747 M		3,137721 M		3,203781 M		86 - 101	5L - 6A W
91	5.085889 W		4,891233 W		4,91714 W		86 - 103	5L - 6A W
92	5.211063 W		5,121536 W		5,32627 W		86 - 111	5L - 7H W
93	5.678361 W	**0,418361**	6,286263 W	**1,226263**	5,699751 W	**0,439751**	86 - 129	5L - 8A W
94	4.049288 W		4,901229 W		4,461738 W		92 - 101	5L - 6A W
95	4.689141 W		5,083341 W		4,899347 W		96 - 101	5L - 6A W
96	2.649137 S		2,844316 M		2,728406 M		101 - 103	6A - 6A M
97	2.733507 M		2,685267 S		2,705436 M		101 - 106	6A - 6A W
98	2.864871 M	**0,464871**	2,608174 S	**0,208174**	2,637734 S	**0,237734**	101 - 111	6A - 7H S
99	2.563008 S		2,552262 S		2,550594 S		103 - 106	6A - 6A W
100	5.021387 W		4,809525 W		4,947478 W		103 - 125	6A - 7H W
101	4.421669 W	**0,131669**	4,540417 W	**0,250417**	4,420745 W	**0,130745**	103 - 129	6A - 8A W
102	3.240624 M		3,384124 W		3,386866 W		106 - 111	6A - 7H W
103	4.500187 W		4,206592 W		4,153284 W		106 - 125	6A - 7H W
104	6.022495 W	**0,142495**	5,786249 W		5,521273 W		106 - 121	6A - 7H W
105	4.502183 W		4,178367 W		4,46182 W		106 - 129	6A - 8A W
106	2.561750 S		2,571669 S		2,532517 S		111 - 113	7H - 7H M
107	2.822090 M		2,730753 M		2,764855 M		111 - 115	7H - 7H M
108	3.882526 W		3,581847 W		3,613526 W		111 - 125	7H - 7H W
109	2.596835 S		2,717903 M		2,685239 S		113 - 115	7H - 7H W
110	2.511877 S		2,531509 S		2,534893 S		113 - 116	7H - 7H M
111	4.051859 W		3,144865 M		3,171786 M		113 - 125	7H - 7H W
112	1.724340 S		1,744684 S		1,743471 S		115 - 116	7H - 7H S
113	2.884788 M		2,896501 M		2,910107 M		115 - 125	7H - 7H W
114	3.115112 M		3,649993 W		3,610001 W		115 - 129	7H - 8A W
115	2.875020 M		3,037596 M		3,010439 M		116 - 125	7H - 7H W
116	4.244176 W		4,250465 W		4,249723 W		125 - 121	7H - 7H W
117	5.492454 W		6,163555 W	**0,163555**	6,499431 W	**0,499431**	121 - 166	7H - 10A W
118	2.728131 M		2,775774 M		2,753066 M		129 - 131	8A - 8A M
119	2.766011 M		2,730896 M		2,738505 M		129 - 134	8A - 8A W
120	3.125551 M		3,023294 W		2,932603 M		129 - 139	8A - 9K M
121	2.569694 S		2,552288 S		2,552064 S		131 - 134	8A - 8A W
122	6.337746 W	**0,737746**	6,704355 W	**1,104355**	6,804648 W	**1,204648**	131 - 166	8A - 10A W
123	2.926513 M		3,068404 M		3,088769 M		134 - 139	8A - 9K W

124	2.694155 S	**0,174155**	2,712211 M	**0,192211**	2,820182 M	**0,300182**	139 - 141	9K - 9K S
125	2.603026 S		2,657329 S		2,632627 S		139 - 143	9K - 9K M
126	2.876994 M		2,914786 M		2,880717 M		139 - 144	9K - 9K W
127	3.273612 M		2,883589 M		2,893977 M		139 - 146	9K - 9K W
128	2.668975 S		2,628776 S		2,614102 S		141 - 143	9K - 9K M
129	2.635522 S		2,553453 S		2,570696 S		141 - 144	9K - 9K S
130	2.690886 S		2,750388 M		2,770452 M		141 - 146	9K - 9K W
131	2.836657 M		2,858819 M		2,84865 M		143 - 146	9K - 9K W
132	3.486357 W		3,161747 M		3,151728 M		143 - 161	9K - 10A W
133	2.503484 S		2,486484 S		2,484928 S		144 - 146	9K - 9K S
134	2.886292 M		3,014051 M		3,022923 M		144 - 161	9K - 10A M
135	2.712177 M		2,782918 M		2,783086 M		146 - 149	9K - 9K W
136	2.786234 M		2,753629 M		2,761243 M		161 - 166	10A - 10A W
137	2.572003 S		2,554823 S		2,553761 S		163 - 166	10A - 10A W
138	2.622237 S		2,617021 S		2,60473 S		171 - 173	11G - 11G W

The average of the upper bound violation can be found the sum of all upper bound violation values by dividing the total number of proton pairs. The higher the number of violations, the higher the calculated average, the higher the total number of proton pairs. [48] The NOE values marked in yellow in Table 1 have an upper bound violation. These values are used in the calculation of average violation. The average upper bound violation was calculated by running the C program script in **Appendix (A2)**.

Table 2: Average upper bound violation $(r)^{-6}$ and number of violations obtained trajectories of the three different force fields.

AMBER15SB		AMBER99SB-STAR-ILDN		AMBER99-4fs-STAR-ILDN	
Average violation $(r)^{-6}$	Number of violations	Average violation $(r)^{-6}$	Number of violations	Average violation $(r)^{-6}$	Number of violations
0,052120	12	0,046191	12	0,046897	11

According to the calculations, the violation and their average upper bound violation values were 12 and 0.052120 Å for 15SB force field, 12 and 0.046191 Å for STAR force field, and 11 and 0.046897 Å for 4fs-STAR force field respectively as shown Table 2. Violation averages belong to the three force fields were found to be approximately 0.05 Å, indicating that the results are very close to the experimental data. The closest result to the experimental data was found with the STAR force field. STAR and 4fs-STAR force fields have very close average violation values. The average violation value of 15SB force field was found to be above 0.05 Å.

4. DISCUSSION AND CONCLUSION

The aim of the experiment was to validate the AMBER15SB force field by comparing with the trajectory analysis results of other force fields on the MD simulation of an α-Lactalbumin-derived peptide. Firstly, the α-La derived peptide INYWLAHAKAG defined from Araki and Tamura was preferred because the its experimental data were more suitable for comparison. The subject of this thesis is determined from the previous research of Patapati and Glykos and the results of the trajectory analysis will be compared with the previous results. [47]

First of all, simulations with NAMD program using 15SB, STAR and 4fs-STAR force fields were carried out before this study in accordance with standard system preparation and simulation protocols. Simulation times with a final temperature of 320K are 6.2 μs, 2.5 μs and 3.5 μs respectively.

Trajectory analysis was performed with *carma* and *grcarma* has graphical interface programs. RMSD analysis is an MD data analysis aimed at observing their dynamic behavior by calculating the distance of the backbone atoms of the superimposed peptide over atomic positions. RMSD magnitudes are represented by colors are blue, yellow and red respectively from low to high. The RMSD magnitude unit is the Angstrom, the distance unit. The lower the RMSD, the greater the similarity of the two superimposed peptides. RMSD analysis was calculated according to the Cα atoms found in the 101-111 residues of α-La, as shown in Figure 5. The result of the STAR force filed shows that the peptide remains more stable than the others. RMSD magnitude for the 15SB and 4fs-STAR force field showed as medium. This means that the molecule does not remain stable for a long time.

The STRIDE method for secondary structure analysis creates an algorithm using dihedral angle and hydrogen bond energies, which enables the analysis of the tertiary structure. When interpreting the secondary structure analysis, the comparison of the results with RMSD analysis facilitated the analysis of concurrent data as shown in Figure 9. In the secondary structure analysis, each secondary structure is expressed in a different color. In the RMSD analysis, the areas with low RMSD magnitude correspond to the helix structures from the secondary structures. In this case, the helix structures provide stability for the α-La derived peptide. The β -turn structures correspond to the areas where the RMSD magnitude is medium, while the β -sheet structure indicates α-La derived-peptide is unstable. Secondary structure analysis also shows the change in time of each residue. Accordingly, it was found that helix structures between 103-107 residues -YWLAH- were common in all three trajectories. Random coil and

β -turn structures are observed in the C terminal of the A-La derived peptide. According to experimental data, 3_{10}-helix structure is observed between 103-106 residues and disordered structure is observed in C terminal toward 111st residue that proves the accuracy of our analysis.

dPCA analysis enables low-dimensional free energy landscapes to be obtained from high-dimensional data. dPCA allows conformational dynamic interpretation of small peptides using dihedral angles (φi, ψi) of the backbone atoms of the peptide. Two-dimensional landscapes were obtained by dPCA analysis by *grcarma* analysis program. The view of landscapes from three different angles is given in Figure13. In the diagram, the blue regions are known to be various energy minima. Accordingly, the unsharpened and unclear blue zone is due to their uncertain energy minima bounder. The inability to maintain stability of the peptide seen in the RMSD analysis of 15SB, STAR and 4fs-STAR force fields can be explained by the large and scattered blue regions herein. Due to the excess of local minima, the presence of unstable intermediate states similar to the stable structure can be observed.

Quantitative comparison of experimental data with trajectory analysis results is the most important step for validation. Nuclear Overhauser effect was used for this study from NMR observables. NOE is defined as the spin population of one nuclei which is affected by the radiation of another nuclei. NOE relates to the inter-atomic distance $(r)^{-6}$ between these two nuclei. Experimental classification is made according to distance amounts: Strong for 1.8-2.7 Å, medium for 2.7 - 3.3 Å, weak for 3.3-5.0 Å. NOE above 5.5 Å is not evaluated. There are 178 atoms in the 101-111 residues of the α-La derived peptide, with a total of 138 NOEs determined. Compared with $(r)^{-6}$ of experimental data, upper bound violations were detected and calculated and the yellow rank is given in Table 1. The violation was calculated by subtracting the $(r)^{-6}$ value from the trajectory analysis from the experimental data's $(r)^{-6}$ value. The sum of obtained violation values were divided by the total number of proton pairs therefore, average violation $(r)^{-6}$ in Table 2 were obtained.

Previous results containing NOE values of α-La derived-peptide prepared by Patapati and Glykos are shared in Table 3. According to their study, NOE violation was calculated to compare the experimental data from trajectories obtained from simulation performed with the CHARMM, AMBER99SB and OPLS force fields.

Table 3: The $(r)^{-6}$ values used in NOE analysis to validate the simulations prepared with three different force fields, CHARMM, AMBER99SB, and OPLS, calculated by Patapati and Glykos in the previous studies with experimental data, combined with the NOE analysis found in Table 1. [37] Yellow color indicates NOE values with upper bound violation.

Pair	CHARMM $(r)^{-6}$	AMBER $(r)^{-6}$	OPLS $(r)^{-6}$	15SB $(r)^{-6}$	STAR $(r)^{-6}$	4fs-STAR $(r)^{-6}$	Proton number	Residue number & exp. classification
1	2,476877 S	2,462987 S	2,426829 S	2.481071 S	2,462541 S	2,461213 S	3 - 6	1I - 1I W
2	2,484441 S	2,466621 S	2,464868 S	2.583273 S	2,457837 S	2,461241 S	6 - 8	1I - 1I M
3	2,87736 M	2,948223 M	2,898257 M	2.956636 M	2,917884 M	2,874216 M	6 - 11	1I - 1I W
4	3,373645 W	3,246874 M	4,238819 W	3.079931 M	3,330569 W	3,362083 W	6 - 18	1I - 1I W
5	2,581993 S	2,549525 S	2,546373 S	2.546668 S	2,548021 S	2,545642 S	8 - 11	1I - 1I W
6	2,601556 S	2,760787 M	1,251396 S	2.685134 S	2,789152 M	2,78731 M	8 - 15	1I - 1I M
7	2,492522 S	2,905752 M	2,486128 S	3.268227 M	2,835574 M	2,798298 M	8 - 23	1I - 2N W
8	3,192265 M	3,198301 M	3,175068 M	3.149422 M	3,176046 M	3,184574 M	11 - 18	1I - 1I W
9	3,99577 W	3,933948 W	4,225215 W	3.537251 W	4,005717 W	3,969294 W	11 - 23	1I - 2N W
10	5,735552 W	5,636932 W	5,343217 W	4.523066 W	6,068139 W	5,91945 W	11 - 45	1I - 3Y W
11	5,151201 W	4,971827 W	4,689839 W	4.699332 W	5,169561 W	5,255258 W	11 - 47	1I - 3Y W
12	1,776638 S	1,750836 S	1,725992 S	1.734409 S	1,751424 S	1,749962 S	14 - 15	1I - 1I S
13	2,597817 S	2,548548 S	2,583904 S	2.538267 S	2,548755 S	2,545186 S	14 - 18	1I - 1I W
14	5,300488 W	5,047209 W	4,139524 W	5.287613 W	5,174417 W	5,167828 W	18 - 47	1I - 3Y W
15	7,380121 W	6,781684 W	4,422224 W	7.394228 W	6,070646 W	6,407717 W	18 - 60	1I - 4W W
16	2,719445 M	2,865915 M	2,920366 M	2.839251 M	2,838034 M	2,872849 M	23 - 25	2N - 2N W
17	2,693992 S	2,587309 S	2,384034 S	2.560532 S	2,528912 S	2,517562 S	23 - 27	2N - 2N W
18	2,64459 S	2,461383 S	2,520366 S	2.802993 M	2,626269 S	2,655377 S	23 - 28	2N - 2N W
19	3,585996 W	3,575119 W	3,024342 M	3.203585 M	3,349983 W	3,558538 W	23 - 37	2N - 3Y W
20	2,547339 S	2,447474 S	2,505926 S	2.612762 S	2,575236 S	2,590415 S	25 - 27	2N - 2N S
21	2,552376 S	2,925984 M	3,048619 M	2.595047 S	2,691438 S	2,675681 S	25 - 28	2N - 2N M
22	4,31112 W	4,187457 W	4,735712 W	4.178218 W	4,335325 W	4,279774 W	25 - 58	2N - 4W W
23	1,785481 S	1,752019 S	1,730807 S	1.740596 S	1,752025 S	1,751077 S	27 - 28	2N - 2N S
24	2,821589 M	2,453421 S	2,626861 S	3.420220 W	2,464163 S	2,47061 S	27 - 33	2N - 2N W
25	2,882844 M	3,592844 W	4,121931 W	3.089009 M	3,641782 W	3,60783 W	27 - 37	2N - 3Y W
26	3,809304 W	4,782897 W	6,96907 W	4.167309 W	3,986824 W	3,93695 W	27 - 86	2N - 5L W
27	2,772072 M	3,720077 W	3,609183 W	2.563088 S	3,756933 W	3,745095 M	28 - 32	2N - 2N W
28	2,766785 M	2,619558 S	0,912858 S	3.542161 W	2,675382 S	2,657931 S	28 - 33	2N - 2N W
29	3,327892 M	3,802367 W	3,154815 M	3.186904 M	3,11418 M	3,072933 M	28 - 37	2N - 3Y W
30	1,732771 S	1,749968 S	1,734492 S	1.735827 S	1,749397 S	1,749062 S	32 - 33	2N - 2N S
31	4,89153 W	3,47432 W	4,066269 W	5.619146 W	4,128271 W	4,127332 W	32 - 62	2N - 4W W
32	2,766871 M	2,841336 M	2,90271 M	2.818651 M	2,811838 M	2,807048 M	37 - 39	3Y - 3Y M
33	2,365418 S	2,538474 S	2,639386 S	2.903367 M	2,582184 S	2,606757 S	37 - 41	3Y - 3Y W
34	3,328684 M	3,241333 M	3,13725 M	2.903319 M	3,027708 M	3,04339 M	37 - 45	3Y - 3Y W
35	2,771402 M	2,617637 S	2,982893 M	2.607674 S	2,751451 M	2,644531 S	37 - 58	3Y - 4W S
36	2,654253 S	2,514538 S	2,677584 S	2.480405 S	2,484259 S	2,483744 S	39 - 41	3Y - 3Y W
37	2,737681 M	3,055685 M	3,051029 M	3.341477 W	3,115395 M	3,117915 M	39 - 45	3Y - 3Y W
38	5,01915 W	5,333897 W	5,333804 W	5.659331 W	5,424614 W	5,442091 W	39 - 47	3Y - 3Y W
39	3,439156 W	3,763442 W	4,926066 W	3.755292 W	3,588957 W	3,641154 W	39 - 101	3Y - 6A W

40	2,936101 M	3,38476 W	4,621804 W	3.418155 W	3,165843 M	3,233196 M	39 - 106	3Y - 6A W
41	2,658835 S	2,779489 M	2,717386 M	2.766231 M	2,75336 M	2,762494 S	41 - 45	3Y - 3Y W
42	4,941253 W	5,041003 W	4,984722 W	5.000275 W	5,01851 W	5,022608 W	41 - 47	3Y - 3Y W
43	3,220603 M	3,583274 W	3,035159 M	3.097927 M	3,53509 M	3,50161 W	41 - 58	3Y - 4W W
44	4,934042 W	5,178456 W	4,706809 W	4.812158 W	5,042621 W	5,034573 W	41 - 60	3Y - 4W W
45	2,458421 S	2,473482 S	2,449224 S	2.454283 S	2,472979 S	2,471438 S	45 - 47	3Y - 3Y W
46	4,300821 W	3,863489 W	3,912751 W	3.678810 W	3,778117 W	3,841277 W	45 - 58	3Y - 4W W
47	4,514317 W	4,15058 W	5,105768 W	4.006906 W	4,106526 W	4,17221 W	45 - 60	3Y - 4W W
48	5,264848 W	4,976714 W	5,5305 W	4.954850 W	5,103781 W	5,232162 W	45 - 62	3Y - 4W W
49	6,034871 W	4,368136 W	7,929763 W	4.240334 W	4,346666 W	4,262186 W	45 - 77	3Y - 4W W
50	5,785583 W	5,328161 W	7,140957 W	6.236425 W	5,738873 W	5,721197 W	45 - 73	3Y - 4W W
51	6,340058 W	4,070195 W	7,282175 W	3.813563 W	3,938153 W	3,83725 W	47 - 75	3Y - 4W W
52	8,957869 W	4,779252 W	4,562979 W	4.307263 W	5,03412 W	4,480706 W	47 - 86	3Y - 5L W
53	6,096661 W	6,495347 W	5,862594 W	6.043411 W	5,951668 W	6,209247 W	47 - 106	3Y - 6A W
54	4,272725 W	4,958648 W	5,64199 W	5.646613 W	5,134753 W	5,117577 W	47 - 121	3Y - 7H W
55	2,807857 M	2,868432 M	2,907656 M	2.868849 M	2,881896 M	2,888709 M	58 - 60	4W - 4W M
56	2,443501 S	2,641185 S	2,772718 M	2.908583 M	2,693594 S	2,744525 M	58 - 62	4W - 4W M
57	2,643996 S	2,814407 M	2,612621 S	2.654615 S	2,694798 S	2,730372 M	58 - 63	4W - 4W M
58	6,082743 W	3,262445 M	6,49692 W	2.703130 M	2,884226 M	2,804333 M	58 - 77	4W - 4W W
59	2,757074 M	2,588838 S	2,645861 S	2.480391 S	2,587881 S	2,535512 S	58 - 82	4W - 5L M
60	2,584571 S	2,582188 S	2,472563 S	2.469619 S	2,530664 S	2,520886 S	60 - 62	4W - 4W M
61	2,720309 M	2,531801 S	2,726903 M	2.507130 S	2,504976 S	2,474096 S	60 - 63	4W - 4W M
62	3,007054 M	3,189908 M	3,09099 M	3.508286 W	3,251679 M	3,330053 W	60 - 66	4W - 4W W
63	6,039514 W	2,867475 M	6,103509 W	3.323182 W	2,969154 M	3,014963 M	60 - 77	4W - 4W W
64	4,50923 W	4,163689 W	4,802129 W	4.308291 W	4,165286 W	4,19216 W	60 - 101	4W - 6A W
65	3,192169 M	3,706108 W	5,775085 W	4.148744 W	3,934016 W	3,886072 W	60 - 115	4W - 7H M
66	1,779063 S	1,744381 S	1,721904 S	1.727354 S	1,743049 S	1,741325 S	62 - 63	4W - 4W S
67	2,932311 M	2,795803 M	2,924449 M	2.806634 M	2,784751 M	2,799047 M	62 - 66	4W - 4W W
68	5,659861 W	3,047176 M	5,721872 W	3.155252 M	3,130289 M	3,077625 M	62 - 77	4W - 4W W
69	3,155726 M	3,372316 W	3,178532 M	3.413223 W	3,412261 W	3,389986 W	62 - 82	4W - 5L W
70	3,226754 M	3,013712 M	2,84144 M	3.068120 M	3,03979 M	3,030337 M	63 - 66	4W - 4W W
71	2,772314 M	3,367305 W	3,044245 M	3.312272 W	3,246895 M	3,424608 W	63 - 82	4W - 5L W
72	2,570414 S	2,547682 S	2,041124 S	2.559937 S	2,547845 S	2,547413 S	66 - 68	4W - 4W W
73	5,000956 W	5,793813 W	4,908978 W	4.901184 W	5,291152 W	5,035814 W	66 - 92	4W - 5L W
74	1,402208 S	2,469497 S	1,405079 S	2.452209 S	2,469274 S	2,469139 S	77 - 75	4W - 4W M
75	4,439386 W	2,863654 M	3,917152 W	2.886166 M	2,863754 M	2,863458 M	68 - 71	4W - 4W W
76	1,401506 S	2,484941 S	1,408704 S	2.467654 S	2,484915 S	2,484299 S	71 - 73	4W - 4W S
77	2,815429 M	2,743128 M	2,821686 M	2.811329 M	2,861316 M	2,864107 M	82 - 84	5L - 5L M
78	2,384997 S	2,566444 S	2,570424 S	2.535769 S	2,473858 S	2,484257 S	82 - 86	5L - 5L W
79	4,550989 W	4,505717 W	3,658901 W	4.386062 W	4,184616 W	4,234179 W	82 - 92	5L - 5L W
80	2,684674 S	2,414792 S	2,706672 M	2.522661 S	2,368415 S	2,385934 S	82 - 101	5L - 6A M
81	4,202587 W	3,780175 W	4,561918 W	3.921565 W	4,091265 W	4,18119 W	82 - 111	5L - 7H W
82	2,58708 S	2,576197 S	2,670751 S	2.651642 S	2,793281 M	2,795216 M	84 - 86	5L - 5L W
83	2,990485 M	2,988235 M	2,860737 M	2.982991 M	2,984914 M	2,993937 M	84 - 89	5L - 5L W
84	3,175362 M	2,794341 M	3,05428 M	2.949377 M	3,202901 M	3,221246 M	84 - 92	5L - 5L W
85	2,747771 M	3,181994 M	3,240202 M	2.992649 M	2,779591 M	2,77437 M	84 - 96	5L - 5L W

86	6,000532 W	5,317393 W	5,073716 W	5.215048 W	5,554625 W	5,446133 W	84 - 106	5L - 6A W
87	4,318586 W	4,108918 W	4,668634 W	4.114622 W	3,992602 W	4,033632 W	84 - 111	5L - 7H W
88	3,016472 M	2,863073 M	3,304636 W	2.809730 M	2,808839 M	2,798771 S	86 - 92	5L - 5L W
89	2,829436 M	2,963261 M	2,792302 M	3.104759 M	3,29539 M	3,312596 W	86 - 96	5L - 5L W
90	3,222159 M	3,504969 W	2,985391 M	3.290747 M	3,137721 M	3,203781 M	86 - 101	5L - 6A W
91	5,136566 W	5,343976 W	5,193496 W	5.085889 W	4,891233 W	4,91714 W	86 - 103	5L - 6A W
92	5,605295 W	5,282066 W	4,7938 W	5.211063 W	5,121536 W	5,32627 W	86 - 111	5L - 7H W
93	5,853856 W	6,276417 W	4,668083 W	5.678361 W	6,286263 W	5,699751 W	86 - 129	5L - 8A W
94	5,23716 W	3,990701 W	4,450751 W	4.049288 W	4,901229 W	4,461738 W	92 - 101	5L - 6A W
95	5,462039 W	4,86058 W	4,109679 W	4.689141 W	5,083341 W	4,899347 W	96 - 101	5L - 6A W
96	2,795927 M	2,61681 S	2,767691 M	2.649137 S	2,844316 M	2,728406 M	101 - 103	6A - 6A M
97	2,586439 S	2,744573 M	2,778174 M	2.733507 M	2,685267 S	2,705436 M	101 - 106	6A - 6A W
98	2,724977 M	2,669924 S	2,832308 M	2.864871 M	2,608174 S	2,637734 S	101 - 111	6A - 7H S
99	2,591802 S	2,549967 S	2,579528 S	2.563008 S	2,552262 S	2,550594 S	103 - 106	6A - 6A W
100	7,775274 W	5,145919 W	8,130695 W	5.021387 W	4,809525 W	4,947478 W	103 - 125	6A - 7H W
101	4,575329 W	4,546557 W	4,709963 W	4.421669 W	4,540417 W	4,420745 W	103 - 129	6A - 8A W
102	3,122429 M	3,445089 W	3,050487 M	3.240624 M	3,384124 W	3,386866 W	106 - 111	6A - 7H W
103	5,897436 W	4,012079 W	8,146884 W	4.500187 W	4,206592 W	4,153284 W	106 - 125	6A - 7H W
104	5,06484 W	5,328755 W	7,986174 W	6.022495 W	5,786249 W	5,521273 W	106 - 121	6A - 7H W
105	5,684398 W	4,756219 W	4,453695 W	4.502183 W	4,178367 W	4,46182 W	106 - 129	6A - 8A W
106	2,867305 M	2,657172 S	2,726426 M	2.561750 S	2,571669 S	2,532517 S	111 - 113	7H - 7H M
107	2,491457 S	2,743007 M	2,789331 M	2.822090 M	2,730753 M	2,764855 M	111 - 115	7H - 7H M
108	5,920891 W	3,601413 W	6,541865 W	3.882526 W	3,581847 W	3,613526 W	111 - 125	7H - 7H W
109	2,693242 S	2,695293 S	2,503014 S	2.596835 S	2,717903 M	2,685239 S	113 - 115	7H - 7H W
110	2,661486 S	2,528262 S	2,838471 M	2.511877 S	2,531509 S	2,534893 S	113 - 116	7H - 7H M
111	4,945799 W	3,197193 M	5,401338 W	4.051859 W	3,144865 M	3,171786 M	113 - 125	7H - 7H W
112	1,782028 S	1,743826 S	1,725048 S	1.724340 S	1,744684 S	1,743471 S	115 - 116	7H - 7H S
113	4,997789 W	2,905119 M	4,886163 W	2.884788 M	2,896501 M	2,910107 M	115 - 125	7H - 7H W
114	2,763862 M	3,405224 W	3,5951 W	3.115112 M	3,649993 W	3,610001 W	115 - 129	7H - 8A W
115	4,956182 W	3,010417 M	2,850413 M	2.875020 M	3,037596 M	3,010439 M	116 - 125	7H - 7H W
116	2,509283 S	4,250422 W	2,513166 S	4.244176 W	4,250465 W	4,249723 W	125 - 121	7H - 7H W
117	7,687438 W	6,375388 W	5,8376 W	5.492454 W	6,163555 W	6,499431 W	121 - 166	7H - 10A W
118	2,773153 M	2,724425 M	2,882388 M	2.728131 M	2,775774 M	2,753066 M	129 - 131	8A - 8A M
119	2,612334 S	2,737517 M	2,757476 M	2.766011 M	2,730896 M	2,738505 M	129 - 134	8A - 8A W
120	2,733574 M	3,057729 M	2,9387 M	3.125551 M	3,023294 W	2,932603 M	129 - 139	8A - 9K M
121	2,590445 S	2,553206 S	2,585576 S	2.569694 S	2,552288 S	2,552064 S	131 - 134	8A - 8A W
122	6,037282 W	6,698218 W	7,233242 W	6.337746 W	6,704355 W	6,804648 W	131 - 166	8A - 10A W
123	3,171054 M	3,140962 M	3,009563 M	2.926513 M	3,068404 M	3,088769 M	134 - 139	8A - 9K W
124	2,694629 S	2,717522 M	2,904556 M	2.694155 S	2,712211 M	2,820182 M	139 - 141	9K - 9K S
125	2,528828 S	2,669834 S	2,652052 S	2.603026 S	2,657329 S	2,632627 S	139 - 143	9K - 9K M
126	2,756244 M	2,938565 M	2,948207 M	2.876994 M	2,914786 M	2,880717 M	139 - 144	9K - 9K W
127	2,875647 M	2,932639 M	2,714568 M	3.273612 M	2,883589 M	2,893977 M	139 - 146	9K - 9K W
128	2,618939 S	2,628628 S	2,609807 S	2.668975 S	2,628776 S	2,614102 S	141 - 143	9K - 9K M
129	2,608879 S	2,558031 S	2,520549 S	2.635522 S	2,553453 S	2,570696 S	141 - 144	9K - 9K S
130	2,791416 M	2,752453 M	2,806003 M	2.690886 S	2,750388 M	2,770452 M	141 - 146	9K - 9K W
131	2,889421 M	2,858272 M	2,954073 M	2.836657 M	2,858819 M	2,84865 M	143 - 146	9K - 9K W

132	3,226134 M	3,164548 M	2,769641 M	3.486357 W	3,161747 M	3,151728 M	143 - 161	9K - 10A W
133	2,517119 S	2,486882 S	2,479317 S	2.503484 S	2,486484 S	2,484928 S	144 - 146	9K - 9K S
134	3,130442 M	3,022848 M	2,970727 M	2.886292 M	3,014051 M	3,022923 M	144 - 161	9K - 10A M
135	2,859677 M	2,777214 M	2,896229 M	2.712177 M	2,782918 M	2,783086 M	146 - 149	9K - 9K W
136	2,750003 M	2,756687 M	2,775639 M	2.786234 M	2,753629 M	2,761243 M	161 - 166	10A - 10A W
137	2,579375 S	2,555016 S	2,582133 S	2.572003 S	2,554823 S	2,553761 S	163 - 166	10A - 10A W
138	3,072851 M	2,60658 S	1,249739 S	2.622237 S	2,617021 S	2,60473 S	171 - 173	11G - 11G W

According to the results, the total number of violations and their averages are given in Table 4. 11 violations have been detected in the AMBER99SB, which is equal to almost half of the violations of other force fields. AMBER99SB average violation $(r)^{-6}$ is 0.050883 Å gave the closest result to experimental data. The average violation results in this study are given in Table 2 and the number of violations varies between 11-12. The lowest average violation $(r)^{-6}$ value is 0.046191 Å was achieved with STAR force field. The average violation $(r)^{-6}$ of the 4fs-STAR force field is with 0.046897 Å almost identical to the STAR force field's result. However, average violation $(r)^{-6}$ value of 15SB force field is 0.052120 Å could not provide the closest result to experimental data, yet a reasonable value was obtained.

Table 4: Average upper bound violation $(r)^{-6}$ and number of violations obtained trajectories of the three different force fields were calculated by Patapati and Glykos in the previous studies.

CHARMM		AMBER99SB		OPLS	
Average violation $(r)^{-6}$	Number of violations	Average violation $(r)^{-6}$	Number of violations	Average violation $(r)^{-6}$	Number of violations
0,137879	21	0,050883	11	0,192171	25

As a result, validation of 15SB force field via MD simulation of α-La derived peptide was achieved using experimental data and STAR and 4fs-STAR force fields. Reasonable results for 15SB were recorded according to experimental data. By comparing the trajectories analysis of the different force fields used in the simulations, it was proved that the closest result to the experimental data was obtained with STAR force field.

REFERENCES

1- Lehninger, A.L., Nelson, D.L., Cox, M.M. (1970). *Principles of Biochemistry* (7nd ed.). Macmillan Learning

2- Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. *Science*. 181 (4096): 223–230. doi:10.1126/science.181.4096.223.

3- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walters, P. (2002). *Molecular Biology of the Cell:The Shape and Structure of Proteins* (4nd ed.). New York and London: Garland Science. ISBN 978-0-8153-3218-3.

4- Anfinsen, C.B. (July 1972). The formation and stabilization of protein structure. *The Biochemical Journal*. 128 (4): 737–49. doi:10.1042/bj1280737.

5- Kim, P.S., & Baldwin, R.L. (1990). Intermediates in the folding reactions of small proteins. *Annual Review of Biochemistry*. 59: 631–60. doi:10.1146/annurev.bi.59.070190.003215.

6- Kubelka, J., Hofrichter, J., Eaton, W.A. (2004). The protein folding 'speed limit'. *Current Opinion in Structural Biology*. 14 (1): 76–88. doi:10.1016/j.sbi.2004.01.013.

7- Zwanzig, R., Szabo, A., Bagchi, B. (1992). Levinthal's paradox. *Proceedings of the National Academy of Sciences of the United States of America.* 89(1): 20–22. doi: 10.1073/pnas.89.1.20

8- Bryngelson, J.D., Onuchic, J.N., Socci, N.D., Wolynes, P.G. (1995). Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins.* 21 (3): 167–95. doi:10.1002/prot.340210302.

9- Leopold, P.E., Montal. M., Onuchic, J.N. (1992). Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proceedings of the National Academy of Sciences of the United States of America*. 89 (18): 8721–5. doi:10.1073/pnas.89.18.8721.

10- Honig, B., & Yang, A. (1995). Free Energy Balance In Protein Folding. *Advances in Protein Chemistry* 46: 27-58. Print.

11- Bryngelson, J. D. et al. Funnels, Pathways, And The Energy Landscape Of Protein Folding: A Synthesis. *Proteins: Structure, Function, and Genetics*. 21(3): 167-195. Print.

12- Rizzuti, B., Daggett, V. (2013). Using simulations to provide the framework for experimental protein folding studies. *Archives of Biochemistry and Biophysics*. 531 (1–2): 128–35. doi:10.1016/j.abb.2012.12.015.

13- Schaefer, M., Bartels, C., Karplus, M. (1998). Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model. *Journal of Molecular Biology*. 284 (3): 835–48. doi:10.1006/jmbi.1998.2172

14- Ruhong, Z. et al. (2004). Hydrophobic Collapse In Multi-domain Protein Folding. *Science* 305.5690 :1605-1609. Print.

15- Fersht, A. R. (1995). Optimization of Rates of Protein Folding: The Nucleation-Condensation Mechanism and Its Implications. *Proceedings of the National Academy of Sciences of the United States of America*. 92.24 :10869-10873. Print

16- Karplus, M., & Weaver, D. L. (1994). Protein Folding Dynamics: The Diffusion-Collision Model and Experimental Data. *Protein Science: A Publication of the Protein Society* 3.4 : 650–668. Print

17- Drenth, J. (2007). *Principles of Protein X-Ray Crystallography*. Springer Science & Business Media. ISBN 978-0-387-33746-3.

18- Royer, C.A. (2006). Probing protein folding and conformational transitions with fluorescence. *Chemical Reviews*. 106 (5): 1769–84. doi:10.1021/cr0404390.

19- Beatrice, M.P., Huyghues-Despointes, C., Pace, N., Englander, S. W., and Scholtz, J. M. (2001). Measuring the Conformational Stability of a Protein by Hydrogen Exchange. *Methods in Molecular Biology*. Kenneth P. Murphy Ed. Humana Press, Totowa, New Jersey, pp. 69–92

20- Bedouelle, H. (2016). Principles and equations for measuring and interpreting protein stability: From monomer to tetramer. *Biochimie*. 121: 29–37. doi:10.1016/j.biochi.2015.11.013.

21- Alder, B. J., Wainwright, T. E. (1959). Studies in Molecular Dynamics. I. General Method. *The Journal of Chemical Physics* 31 (2): 459. doi:10.1063/1.1730376.

22- EMBnet.ch. (2018). Home MD. [online] Available at: https://embnet.vital-it.ch/MD_tutorial/ [Accessed 20 May 2019].

23- Mackerell, A. D. (2004). Empirical force fields for biological macromolecules: overview and issues. *Journal of Computational Chemistry*. 25(13):1584-604 doi: 10.1002/jcc.20082

24- Ambermd.org. (2019). The AMBER (Assisted Model Building with Energy Refinement) Molecular Dynamics Package. [online] Available at: http://ambermd.org [Accessed 25 May 2019].

25- Charmm.org. (2019). CHARMM (Chemistry at Harvard Macromolecular Mechanics) Molecular Dynamics Package. [online] Available at: https://www.charmm.org/charmm/?CFID=03380c6f-266a-494d-bc3d-4741256b2f84&CFTOKEN=0 [Accessed 25 May 2019].

26- Gromos.net. (2019) GROMOS (Groningen Molecular Simulation) Molecular Dynamics Package. [online] Available at: http://www.gromos.net/ [Accessed 25 May 2019].

27- .Bizzarri, A.R. & Cannistraro, S. (2017). Molecular Dynamics of Water at The Protein−Solvent Interface. *The Journal of Physical Chemistry* B 106.26: 6617-6633.

28- Jacopo, T., Mennucci, B., Cammi, R. (2005). Quantum Mechanical Continuum Solvation Models. *Chemical Reviews* 105.8 (2005): 2999-3094.

29- Hall, L., Emery, D.C., Davies, M.S., Parker, D., Craig, R.K. (1987).Organization and sequence of the human alpha-lactalbumin gene". *Biochem. J.* 242 (3): 735–42. PMC 1147772. PMID 2954544.

30- Qasba, P.K., Kumar, S. (1997). Molecular divergence of lysozymes and alpha-lactalbumin. *Critical Reviews in Biochemistry and Molecular Biology*. 32 (4): 255–306. doi:10.3109/10409239709082574. PMID 9307874

31- Demarest, S.J., Hua, Y., Raleig, D.P. (1999). Local Interactions Drive the Formation of Nonnative Structure in the Denatured State of Human R-Lactalbumin: A High Resolution Structural Characterization of a Peptide Model in Aqueous Solution. *Biochemistry*. 38. 7380-7387. Print.

32- Araki, M. & Tamura, A. (2007). Transformation of an α -Helix Peptide Into a β -Hairpin Induced by Addition of a Fragment Results in Creation of a Coexisting State. *Proteins: Structure, Function, and Bioinformatics* 66:860–868. Print.

33- Serafeim, A.P., Salamanos, G., Patapati, K.K., Glykos, N.M. (2016). Sensitivity of Folding Molecular Dynamics Simulations to Even Minor Force Field Changes, *Journal of Chemical Information and Modeling.* doi: 10.1021/acs.jcim.6b00493

34- Gromacs.org (2019). GROMACS (Groningen Molecular Simulation) Molecular Dynamics Package. Force Fields [online] Available at: http://www.gromacs.org/Downloads/User_contributions/Force_fields [Accessed 21 June 2019].

35- Hopkins, C. W., Grand S.L., Walker, R.C., Roitberg, A.E. (2015). Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning, *Journal of Chemical Theory and Computation*, 11, 1864−1874, doi: 10.1021/ct5010406

36- Kresten, L.L., et al. (2010).bImproved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field, *Proteins*, 78.8: 1950–1958.

37- Glykos, N. M. (2006). Software News And Updates Carma: A Molecular Dynamics Analysis Program, *Journal of Computational Chemistry*, 27.14: 1765-1768. Web. 21 June 2019.

38- Koukos, P.I. & Glykos, N.M., (2013). Grcarma: A Fully Automated Task-Oriented Interface For The Analysis Of Molecular Dynamics Trajectories. *Journal of Computational Chemistry*, 34.26: 2310-2312

39- PSF Files. [online] Available at: https://www.ks.uiuc.edu/Training/Tutorials/namd/namd-tutorial-unix-html/node23.html [Accessed 21 June 2019].

40- Hsin, J. et al. (2008). Using VMD: An Introductory Tutorial. *Current Protocols in Bioinformatics* n. pag. Web. 21 June 2019.

41- Altis, A., Nguyen, P., Hegger, R., and Stock, G. (2007). Dihedral angle principal component analysis of molecular dynamics simulations, *The Journal of Chemical Physics,* 126 (24), p. 244111. doi: 10.1063/1.2746330

42- "Molecular docking, estimating free energies of binding, and AutoDock's semi-empirical force field". Sebastian Raschka's Website. [online] Available at: http://sebastianraschka.com/Articles/2014_autodock_energycomps.html [Accessed 21 June 2019].

43- Bryngelson, J. D. et al. (1995). Funnels, Pathways, And The Energy Landscape of Protein Folding: A Synthesis. *Proteins: Structure, Function, and Genetics* 21.3: 167-195. Print.

44- CARMA(n), A molecular dynamics analysis program. [online] https://utopia.duth.gr/glykos/pdf/carma.pdf Available at: [Accessed 21 June 2019].

45- Frishman, D., Argos, P. (1995). Knowledge-based protein secondary structure assignment. **Proteins** 23(4):566-79. doi:10.1002/prot.340230412 PMID 8749853

46- Hoult, D.I., Bhakar, B. (1997). NMR Signal Reception: Virtual Photons and Coherent Spontaneous Emission, *Concepts Magnetic Resonance*. 9, 277–297.

47- Patapati, K. K. & Glykos, N. M. (2011) Three Force Fields' Views of the 310 Helix, *Biophysical Journal*, 101:1766-1771. doi: 10.1016/j.bpj.2011.08.044

48- Zagrovic, B., and Wilfred F. G. (2006). Comparing Atomistic Simulation Data With The NMR Experiment: How Much Can Noes Actually Tell Us? *Proteins: Structure, Function, and Bioinformatics* 63.1: 210-218.

49- Neuhaus, D., Williamson, M. P. (2000). *The Nuclear Overhauser Effect in Structural and Conformational Analysis, 2nd ed*. Wiley-VCH

50- Mu, Y., Nguyen, P., and Stock, G. (2005). Energy Landscape of a Small Peptide Revealed by Dihedral Angle Principal Component Analysis, *Proteins: Structure, Function, and Bioinformatics*, 58(1), pp.45-52., doi: 10.1002/prot.20310

APPENDIX

A1. Calculation of NOE

```
#include <stdio.h>
#include <math.h>
#include <stdlib.h>

int main(argc,argv)
int     argc;
char    *argv[];
{
  FILE          *in;
  char          c;
  char          string[100];
  int           columns;
  int           junk;
  int           i;
  long double   r6, r3;
  float         val;
  int           nof_values;
  int           nof_values2;
  int           fraction;
  double        *data_r3;
  double        *data_r6;

  double        *sd_data_r3;
  double        *sd_data_r6;


  setlinebuf( stdout );
  in = fopen( argv[1], "r" );
  if ( in == NULL )
    {
      printf("Can't open file %s for reading. Abort.\n", argv[1] );
      exit(1);
    }

  /* How many columns we have on a line ? You won't believe this ... */
  columns = 0;
  while( fscanf(in, "%c", &c) == 1 )
    {
      if ( c == '\n' )
        break;
      if ( c == '.' )
        columns++;


    }

  fprintf(stderr, "Found %d columns with distances.\n", columns );
  fprintf(stderr, "Allocating memory ...");
  data_r3 = (double *)(calloc( columns, sizeof(double)));
  data_r6 = (double *)(calloc( columns, sizeof(double)));
  sd_data_r3 = (double *)(calloc( columns, sizeof(double)));
  sd_data_r6 = (double *)(calloc( columns, sizeof(double)));
  if ( data_r3 == NULL || data_r6 == NULL || sd_data_r3 == NULL || sd_data_r6 == NULL )
    {
      printf("Failed to allocate %ld bytes of memory. Abort.\n", (long int)(2*columns*sizeof(double)));
      exit(1);
    }
  fprintf(stderr, " done.\nNow reading line : ");
  rewind( in );

  nof_values = 0;
  fprintf(stderr, "%8d", nof_values );
  while( 1 )
  {
    if ( fscanf(in, "%s", &string[0]) != 1)     /* skip frame number */
      break;

    for ( i=0 ; i < columns ; i++ )
        if ( fscanf(in, "%f", &val) != 1 )
          {
            printf("Ooops. Corrupt input file (or not from carma ?). Abort.\n");
            exit(1);
          }
        else
              {
                data_r3[i] += 1.0l / (val*val*val);
                data_r6[i] += 1.0l / (val*val*val*val*val*val);
              }

    nof_values++;
    fprintf(stderr, "%8d", nof_values );

  }


  printf("\n\nSecond pass to calculate standard deviations ...\n");


  rewind( in );

  nof_values2 = 0;
  fprintf(stderr, "%8d", nof_values2 );
  while( 1 )
  {
    if ( fscanf(in, "%s", &string[0]) != 1)     /* skip frame number */
      break;
```

```
   for ( i=0 ; i < columns ; i++ )
       if ( fscanf(in, "%f", &val) != 1 )
         {
           printf("Ooops. Corrupt input file (or not from carma ?). Abort.\n");
           exit(1);
         }
       else
            {
              sd_data_r3[i] += ( 1.0l / (val*val*val) - (data_r3[i]/nof_values) ) * ( 1.0l / (val*val*val) - (data_r3[i]/nof_values) );
              sd_data_r6[i] += (1.0l / (val*val*val*val*val*val) - (data_r6[i]/nof_values)) * (1.0l / (val*val*val*val*val*val) - (data_r6[i]/nof_values));
            }

   nof_values2++;
   fprintf(stderr, "[illegible]%8d", nof_values2 );

 }


 if ( nof_values != nof_values2 )
   {
     printf("Different number of values during second pass ?!? Panic.\n");
     exit(1);
   }


 printf("\nColumn    r^-3    sd(r^-3)   r^-6    sd(r^-6)\n");
 for ( i=0 ; i < columns ; i++ )
 {
   printf("% 5d % 10.6f% 10.6f% 10.6f% 10.6f  \n", i+1, (double)(pow((data_r3[i]/nof_values),-1.0l/3.0l)),
                                      (double)(pow(sqrt(sd_data_r3[i]/nof_values2),-1.0l/3.0l)),
                                      (double)(pow((data_r6[i]/nof_values),-1.0l/6.0l)),
                                      (double)(pow(sqrt(sd_data_r6[i]/nof_values2),-1.0l/6.0l)) );
 }

}
```

A2. Calculation of Average of the Upper Bound Violation

```
#include <stdio.h>

main()
{
        float   exp, md;
        float   sum = 0.0;
        int     number = 0;
        int     violations = 0;

        while( scanf("%f %f", &exp, &md) == 2 )
        {
                number++;
                if ( md > exp )
                        {
                                violations++;
                                sum += md - exp;
                        }
        }

        printf("Over %d NOEs, there are %d violations, average is %f\n", number, violations, sum / number );

}
```