ANALYSING THE INVOLVEMENT OF A CU/ZN SUPEROXIDE DISMUTASE MUTATION IN FAMILIAL AMYOTROPHIC LATERAL SCLEROSIS USING MOLECULAR DYNAMICS SIMULATION TECHNIQUE

A MAJOR PROJECT DISSERTATION SUBMITTED

IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE DEGREE

OF

MATER OF TECHNOLOGY

IN

BIOINFORMATICS

SUBMITTED BY

PRODYOT BANERJEE

(2K19/BIO/07)

UNDER THE GUIDANCE OF

DR. YASHA HASIJA

DEPARTMENT OF BIOTECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY

Shahbad Daulatpur, Main Bawana Road,

Delhi- 110042, India

CERTIFICATE

This is to certify that the M.Tech. Major report entitled "**ANALYSING THE INVOLVEMENT OF A CU/ZN SUPEROXIDE DISMUTASE MUTATION IN FAMILIAL AMYOTROPHIC LATERAL SCLEROSIS USING MOLECULAR DYNAMICS SIMULATION TECHNIQUE**" submitted by Prodyot Banerjee (2K19/BIO/07) in partial fulfillment of the requirement for the award of the degree of Master of Technology from Delhi Technological University, is an authentic record of the candidate's own work carried out by him under my guidance. To best of my knowledge this work has not been submitted in part and full for any Degree or Diploma to this University or elsewhere.

Date: 27th June, 2021

27.06.2021

Dr. Yasha Hasija
Department of Biotechnology
Delhi Technological University

Prof. Pravir Kumar
Head of the Department
Department of Biotechnology
Delhi Technological University

DECLARATION

I Prodyot Banerjee (2K19/BIO/07) student of M.Tech Bioinformatics, hereby declare that the project entitled Dissertation titled "**ANALYSING THE INVOLVEMENT OF A CU/ZN SUPEROXIDE DISMUTASE MUTATION IN FAMILIAL AMYOTROPHIC LATERAL SCLEROSIS USING MOLECULAR DYNAMICS SIMULATION TECHNIQUE**" which is submitted by me to Department of Biotechnology, Delhi Technological University, Delhi in the partial fulfillment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed for the basis for the award of any degree, Diploma Associateship, Fellowship or other similar title or recognition.

Date: 27th June, 2021

PRODYOT BANERJEE
(2K19/BIO/07)

DEPARTMENT OF BIOTECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY

ACKNOWLEDGEMENT

At the time of submission of my M.Tech. Dissertation, I would first like to thank GOD for giving me patience, strength, capability, and willpower to complete my work. Apart from our efforts, the success of this project depends largely on the encouragement and guidelines of many others. I, therefore, take this opportunity to express my gratitude to the people who have been instrumental in the successful completion of this project.

My initial thank is addressed to my mentor **Dr. Yasha Hasija**, Department of Biotechnology, Delhi Technological University, who gave me this opportunity to work in a project under her. It was his enigmatic supervision, constant encouragement and expert guidance which have enabled me to complete this work. I humbly seize this opportunity to express my gratitude to him.

I would also like to extend my sincere gratitude to **Prof. Pravin kumar** for providing basic infrastructure and facilities.

I would like to extend my sincere gratitude to **Mr. Rajkumar Chakraborty** for his keen observation and continuous advice. Her thoughtful inputs on the project issues have been invaluable for the productive progress of the project. I extend my thanks to technical staff Mr. Jitender Singh and Mr. C.B. Singh who had been an aid whenever required. Lastly, I wish to extend my thanks to my family and friends who have supported me through the entire process.

PRODYOT BANERJEE

(2K19/BIO/07)

CONTENTS

LIST OF FIGURES

ABSTRACT

Amyotrophic lateral sclerosis (ALS) is a type of neuro-degenerative disorder that causes muscle weakness, many type of disabilities and can eventually lead to death. Out of all the worldwide Amyotrophic lateral sclerosis cases, 5 to 10% are Familial Amyotrophic lateral sclerosis (FALS) cases. The involvement of superoxide dismutase (SOD1) in Familial Amyotrophic Lateral Sclerosis can be found in many published literatures. I tool a missense mutation of SOD1 gene which causes the substitution of Threonine (Thr) for Alanine (Ala) at the 5th position and analysed its involvement in FALS using molecular dynamics simulation technique. To serve the purpose we first designed the mutant protein i.e. mutant Cu/Zn superoxide dismutase (m4B3E) using homology modeling software SWISS MODEL and analysed stability of the same against the wild type protein i.e. Cu/Zn superoxide dismutase (4B3E) by using molecular dynamics simulation as a tool. I analysed the behavior of the mutant protein and wild type protein simultaneously, as the wild type protein is experimentally proven it is conceded stable. So, I compared the mutant protein to the wild type in the conditions similar to the in vivo environment as molecular dynamics simulation mimics the in vivo environment. We analysed the results in terms of RMSD, RMSF and radius of gyration, further visualization of the simulation was done using VMD (Visual Molecular Dynamics) tool.

Keywords- Familial Amyotrophic lateral sclerosis, Cu/Zn superoxide dismutase, molecular dynamics simulation

INTRODUCTION

Amyotrophic lateral sclerosis is a type of neuro-degenerative disorder that causes muscle weakness, several type of disabilities and can eventually lead to death. It is a combination of motor neuron disorder, mostly comprising of lower and upper motor neuron and affects the cranial nerve in the pons and medulla, spinal cord and frontal cortex. The initial publication suggested that sequence variantion present in the Cu/Zn SOD1 gene was the causative agent of ALS[1]. Out of all the worldwide Amyotrophic lateral sclerosis cases, 5 to 10% are Familial Amyotrophic lateral sclerosis (FALS) cases[2]. The research publications on Familial inheritance of Amyotrophic lateral sclerosis (FALS) began back in 1800s but there was not much attention about the same back then. Later on in the year 1993, with the development of molecular genetic techniques it was found that superoxide dismutase (SOD1) is involved in the FALS. Further finding proved that there are many genes involved in the FALS but most common of them was superoxide dismutase (SOD1)[3]. Although FALS is heterogeneous but most cases follow an X-linked inheritance pattern, autosomal dominant and recessive forms. Mutation at many different positions on the SOD1 gene can cause FALS which can be found in the UniProt and Clinvar databases, one of such missense mutation in the 5th position causes the substitution of Threonine (Thr) for Alanine (Ala) was found in two patient from 1 Japanese family [4], The publication provides an insight about the instability of SOD1 gene for the this mutation and further suggests the involvement of this mutation in FALS. The publication also concluded the possibility of SOD1 mutation at 5th position (T for A) causing FALS. I took this mutation as a lead for our study and wanted to check the behavior of mutated Cu-Zn superoxide dismutase enzyme coded by this mutated SOD1 gene in the in vivo conditions using molecular dynamics simulation technique.

In this Project I have designed a mutated Cu-Zn superoxide dismutase enzyme (m4B3E) which has a missense mutation i.e. T replacing A in the 5th position (UniProt ID- VAR_007130). For designing this protein, I used homology modeling tool SWISS-MODEL. Further my aim was to prove the instability of the mutated protein inside the host which is causing dysfunction of the superoxide dismutase enzyme in turn causing FALS disorder. So in order to achieve my aim I focused on the determination of structural stability of this mutated Cu-Zn superoxide dismutase enzyme(m4B3E) against the wild type Cu-Zn superoxide dismutase

enzyme(PDB Structure id- 4B3E) using Molecular Dynamics simulation software GROMACS and following the protein in water protocol provided in the GROMACS Tutorials (http://www.mdtutorials.com>gmx). I used Molecular Dynamics simulation as a tool to provide a virtual insight into the contribution of this mutation of SOD1 gene in causing FALS disorder by accessing the protein structure and function.

ALS which is the abbreviation for amyotrophic lateral sclerosis is also called as Lou Gehrig's disease; it is a neurodegenerative disorder and is progressive in nature which means it worsens with time. It majorly affect the Nerve cells which control the voluntary muscle movement involving activities like talking and walking[5] etc. As the ALS belongs to the motor neuron disorder it causes the degeneration of nerve cell gradually and eventually lead to the Motor neuron death, due to the same motor neuron stop sending neuronal messages from brain to the muscles which eventually weaken the muscles initially and in the later stages it stops function as the brain can no longer initiate the control over voluntary muscles[6]. Most of the ALS and 5 to 10% of FALS disease are caused by either mutation in the SOD1 gene or mutation in the TARBP gene[7].

The first step towards understanding biology is to understand the structure of the macromolecule and its related molecular interaction, as the interaction governs the function of the macromolecule and even influence the structure. Considering the early days of the determination of macromolecular structures back in 50s, we have come a long way searching for new structure as our knowledge and discovery of macromolecules including proteins and nucleic acids is also growing at a fastidious rate day by day. There are numerous new structures added on to the Protein Data Bank (PDB) with each passing year. As of now total number of entries available on PDB are more than 1,70,000. Although the PDB entries started back in 1976, there is an exponential growth in the addition of annual entries ever since and after the year 2017, more than 11000 entries per year can be observed. This gives us an idea of growing knowledge about the macromolecules in biology, but this knowledge is very basic like biological phenomenon involved in membrane transportation, enzyme regulation, ribosomal structure building and so on. The reason behind considering structural knowledge as basic is that, by only knowing the structure we cannot predict the actual behavior of the macromolecule in vivo. It is true that the 3D structure of the macromolecule has helped in designing the docking states and also aided in

structure based drug designing, but it is only one aspect of molecular behavior considering it has to be validated experimentally in wet labs. But the knowledge of dynamics and energetic of a molecule helps us to determine the behavior of the macro molecule in real environmental conditions i.e. in vivo. With this we can achieve the favorable state of the macromolecule. Further, macromolecules can exhibit allosteric conformation in which it can exist in a state of comparable stability or transition state this is where Molecular dynamics simulation comes into play.

Protein undergoes many conformational changes and rearrangements during its functioning, these rearrangement can facilitate in changing the active site conformation which in turn will change the function of the protein or the entity as a whole, with respect to its environment. There are examples of proteins with small conformational changes in their structure without any changes in their Overall fold, this type of small changes are undetectable by the docking algorithms but there are also evidences of proteins having large conformational changes in their structure. Furthermore, the protein dynamics analysis is important in discovering the transit phenomena in the cavities, alongside proteins nucleic acids also show conformational changes but that is not of our interest as far as this experiment is concerned. To understand the conformational changes, it is important to take energetics/forcefield and dynamics of the molecule into consideration which helps in predicting its behavior in the in vivo environment. This conformational ensemble analysis can be achieved by molecular dynamic simulations approach using several parameters which are defined in the later parts of this report.

The use of M D Simulation in this project is to check the stability to the mutant protein i.e. m4B3E in order to justify the fact that the mutation of SOD1 gene at 5^{th} position (T for A) is a strong causative factor of FALS as the mutation makes the protein unstable in the dynamic condition which implies that it works similarly in the in vivo condition. Although this M D Simulation supports the above stated fact in the computational way but it is not necessary that the protein will behave exactly the same in the wet lab conditions or say real life scenario, but it has been proven that the M D Simulation is the closest technique to mimic the in vivo environmental conditions, so keeping this thing in mind I conducted the experiment with M D Simulation as a tool to detect the stability of the designed mutant protein.

Further Considering the situation where not much homology is available for a protein structure prediction, ab-initio methods are used for prediction with tools like I-TASSER and Rosetta but the accuracy of these tools are not very good and reliable so, the predicted structure is considered to be verified using molecular dynamics simulation in order to get nearby the crystal structure or experimental structure but in this case the homology was good and verified it using BLAST, as the homology was good it was better to go with homology modeling as it is more reliable and accurate because of the availability of template structure. For the same reason I chose SWISS MODEL to build my model as it available online and free of cost and it also finds the template by its own. The detailed and stepwise process of the experiment is defined in the method portion of this report and the flowchart view of the same is represented in fig. 5. The detailed study about the Molecular dynamics simulation and Amyotrophic lateral sclerosis disorder with its symptoms, factors involved, treatment can be found in the published literature which is discussed in the next section of this report i.e. review of literature

REVIEW OF LITERATURE

ALS is a neurodegenerative disorder and is progressive in nature mostly affecting the Nerve cells which control the voluntary muscle movement. The most used Diagnostic method for ALS was electromyography[8] but the further development of technology and study of the disease provided more opportunities to assess the disorder by the introduction of Diagnostic criteria like 'El Escorial' in the year 1994[9] and 'Airlie House' criteria in the year 2000[10] which helped in classifying the patients for different drug trials and Research studies and further classified them into several groups of definite and probable classes[11].The incidents of SALS and FALS shows an increasing ratio since the year 1990 and there is also changes in the male to female ratio gradually[12,13].

The symptoms of this disorder includes muscle atrophy with the involvement of lower and upper motor neuron dysfunctioning[14], further it also contributes to pathological syndrome and more complex clinical features which were defined in the published literature[15-20]. The most common initial signal of this disorder includes muscle twitching, specially the involuntary muscles are involved after which cramps can be observed with weakness in muscles and gradual development of dysfunctioning atrophic limbs, sensory organs, multiple organ disorders and finally leads to parkinsonism featuring dementia. In some patients respiratory symptoms are more prevalent than the limb symptoms[21,22].

There are many factors involved in this neurodegenerative disorder which includes genetic factors[23], oxidative stress[24-27], neurofilament and protein aggregation[28-31] and many more. The targeted treatment of ALS or SALS or FALS is not present at the moment, although many studies were done in the past and is still going on to achieve the same. The proper management of the patient's health is the only way to extend the tenure of life for these patients, this managements include symptomatic treatments[6] (like pain, depression, cramp and anxiety management), nutritional diet management[32,33] and in critical cases (especially with the individual having respiratory symptoms) there is a ventilatory management[34-39].

Molecular Dynamics simulation also called as M D Simulation is not a very old science procedure, considering the fact its development has not even touched 100 years. In the year

1964, Aneesur Rahman who is widely recognized as the 'father of molecular dynamics' used the Newtonian motion equation in order to perform dynamics simulation of a liquid system[40] and later on, Watts and Barker came with a different approach for simulation of water[41] called as Monte Carlo. In the late 70s in the MD simulations of proteins were carried out[42]. After that simulation of other biomolecules were done using free energy calculations[43], force field calculation and protein-ligand docking calculations[44] etc. Since then, the technique has moved a long way and with the advancement of computers this computer intensive technique has reached new heights and is expected to grow further in the coming years. Molecular dynamics simulation approach is accepted worldwide because of the simplicity of its algorithm. New and advanced strategies are being developed nowadays to carry out long MD simulations and to calculate the dynamics of the protein and nucleic acid with high atomic number. During the initial days simulations were carried out for several hundred of atoms but with its course of development now simulation can be carried out for more than 10,00,000 of atoms. With the development high efficiency computing systems having high end processors, advanced graphics processing units (GPUs), multi-core CPU and high memory it is now possible to run MD simulation in microsecond instead of nanoseconds[45].

There are many force fields used in M D simulation and mainly are of two types i.e. additive or not additive but the more generalized is additive. Force field is used to predict molecular enthalpies, vibration spectra and structures. There are various force fields for biological systems[46], OPLS[47], ReaxFF[48], AMOEBA[49], GROMAS[50], CHARMM[51], AMBER[52] are popularly used force fields in MD simulations. Futher the system preparation in which 3D structure generated from experimental techniques or homology modeling or ab initio modeling is taken and its energy minimization is achieved using conjugate-gradient and hybrid steepest-descent technique. A periodic box is setup to cover the minimized structure followed by addition water molecule to the periodic box. Initially MD simulation is run for a short period of time with only water molecules after which the full molecular system having both water and biomolecule is run for MD simulation[53]. MD simulations algorithm has iterations of numerical integration to calculate atomic movements along with conserved energy present within the system by using Newtonian mechanics equation. Further an ensemble is used to adjust within the system. Several ensembles are used in MD simulations including canonical/NVT ensemble, isobaric-

isoenthalpic/NPH ensemble, isobaric-isothermal/NPT ensemble, grand canonical/μVT ensemble, microcanonical/ NVE ensemble. Here, (N) stands for fixed number of atoms , (P) stands for fixed pressure, (T) stands for fixed temperature, (V) for fixed volume , (H) for fixed enthalpy, (μ) represents fixed chemical potential and (E) represents fixed energy of the system[53].

A number of popular software packages for MD simulations are GROMACS[54], AMBER[55], NAMD[56], CHARMM[57], DESMOND[58], TINKER[59], DL_POLY[60] out of all this packages GROMCS is the most widely used as it is its license in open source. M D simulation mimics the actual environmental conditions inside the body of the host so it is used in order to optimize protein models designed using homology modeling[61] and also to determine the stability of the protein. With the advancement of computational capabilities, now it's promising to perform high-end MD simulations in biological systems. The software packages widely used to perform MD simulations have different functions and different ways of calculating MD simulation trajectory. The most interesting part is, every software package has its own way of solving the problem and each package has some advantage over the other. This different software packages can use different force field and are compatible with advanced interfaces, one of which is Message Passing Interface (MPI)[62] that aids in use of multiple computer core at the same time in order to trim down computational time plus making the simulation process more time efficient than before.

A modern day addition to the technique includes division of different segments of simulation among processors in order to make the process faster, this strategy is often regarded as spatial decomposition. With the advancement of GPUs and high performance processor, MD simulation core performance has also improved. ACEMD[63] is MD simulation software packages specifically developed to be used on GPUs. Further, the combination of GPUs and MPI can be employed to perform high throughput MD simulation. Molecular dynamic simulation can be performed for 10-100 nanoseconds with water molecule as solvent to detect the environment[64]. The main advantage of molecular dynamics simulation over classic technique is that it is less time consuming as compared to the experimental wet lab techniques and it also shows the activity of the macromolecules inside the body, like the tension on the macromolecules in actual in vivo environment, the behavior of the macromolecule in the presence of solvent or any charged molecule like sodium, potassium and likes.

We explore one M D Simulation application in this project, although Molecular Dynamic simulation finds its application in many fields of science including chemistry, physics and biology. In the recent years with the advancements in the computational biological techniques, it is now possible to design a drug with the help of computer which is called as Computer Aided Drug Designing (CADD). The CADD has its advantage over traditional techniques like high throughput screening and in fact it can produce high hit rate in case of novel drug discovery. MD simulation is used to validate the newly predicted proteins structure built using either homology modeling[64] or ab initio modeling[65] and furthermore, MD simulation techniques can be combined with traditional experimental structure visualization techniques such NMR and FRET to increase its resolution and quality as discussed earlier[53].Molecular dynamics simulation can be used to solve complex drug discovery issues, for instance issues related with protein ligand interaction stabilities[66-68], process involved in binding[69-71] and the kinetics of binding[72-74]. Further protein ligand interactions can be understood in real life environment or in vivo conditions with the aid of M D simulation. Within few recent years, researchers have discovered the application of molecular dynamics in in various medicinal chemistry fields and use it as a tool for druggable site challenges[75].MD simulation is used for refinement of predicted structures either it be a homology modeled structure or ab initio predicted structure[62].

MD simulation is a confirmatory step in the refinement of models[76] having new amino acid sequences. Simulation also ensures the quality, stability and dynamic balance of the designed model in the in vivo environment. Allostery of the protein can be explored using MD simulation, allostery is use for the study of conformational changes happening in protein at different environmental conditions such as the existence of protein in tensed and relaxed state inside the body. Allostery explains the rearrangement of protein structure in a timely period, collectively called as conformational change[77]. Monod model is one of such model which explains the conformational shift. Understanding the conformational shift is crucial in order to understand the protein's structure as well as function inside the body, because the conformational changes can influence the protein function. So as to understand and analyze the conformational shift efficiently, short simulation can be run instead of large simulation experiments[78]. With the advancements in computational technology many science fields have gained speed and accuracy in their analysis, two of such fields are nanotechnology and toxicology. Toxicology prediction has now become easier with the help of computational techniques[79] and it is also used in drug

designing[80]. Toxicological analyses are important in order to test the toxicity of drugs inside host's body. Normally toxicology measurements take time but it is important to consider these measurements while analyzing newly designed drugs. Nowadays with the help of computational toxicology analysis, the prediction of these measurements has become a lot easier and faster. Further in the recent times, MD simulation is integrated with computational toxicology analysis for detailed study in the field of toxicology[53].

Prediction of peptide structure is also done using molecular dynamics simulation and is reliable source for it as a molecular dynamics simulation consider the configuration entropy and solvent effect which are important to for determining peptide conformations[81,82]. Peptide act as a bridging link between proteins and drug molecule[83] so, right prediction of peptide structure aids in determination of its function and inturn designing of accurate peptide based drugs. The structure of peptide predicted by REMD simulations using implicit solvent[84,85] had high resemblance to the native structure[84] as explained by Dill and co-workers. Further, the cyclic peptide are considered better in structure prediction as compared to the linear peptides as the linear peptide cyclization source very low confirmation flexibility[86,87] REMD simulation in combination with quantum mechanical refinement using GAFF i.e. Generalized AMBER Force Field[88] along with implicit solvation has precisely predicted the X-ray crystallographic structure of cyclic peptoid nonamer as well as N-aryl peptoid trimer with high accuracy[89].

The model refinement and ab initio method of structure prediction has a common issue associated with low knowledge of conformational changes and selection of inappropriate force field, this problem can be resolved using limited and accurate amount of structural information so as to increase the precision of the molecular dynamically predicted structure and inturn molecular dynamics simulation. A limited supply of structural knowledge will take less time and also lower the cost of the experiment, bioinformatic tools have also been used with the aim of predicting secondary structure using data assisted modeling for the after the year 2014, when the eleventh experiment on Critical Assessment for protein Structure Prediction (CASPl1) was initiated data-assisted modeling became a vital part of the CASP experiment[90]. Use of experimentally known protein data for MD based structure prediction is not new it has been in use since 1989, when "Associated Memory Hamiltonians" (AMH) was proposed[91], the structural data obtained from homologous protein is use for learning and creating a memory set

which is helpful in the prediction of new structure which nearly resembles the native structure and all this is achieved by molecular dynamics based simulation annealing[92].

METHOD

1. The first step is preparation of protein structure for MD simulation in both the cases of wild type and mutant protein. The wild type protein structure was available in Protein Data Bank, so we extracted it directly from there in .pdb format, PDB id of the wild type protein was 4B3E i.e. structure of copper-zinc superoxide dismutase complexed with bicarbonate (fig.1). The structure 4B3E was having 8 chains, out of which A and F were in close proximity so we took A chain and F chain for our study.

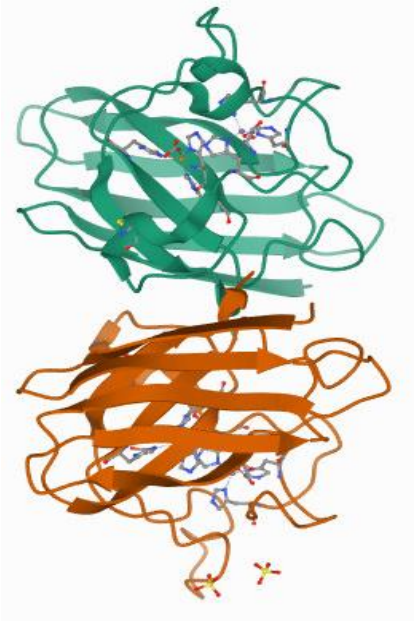

Fig.1. Structure of 4B3E

The structure of mutant protein or say mutant Cu-Zn superoxide dismutase enzyme was not available in the PDB. So, we started designing the mutant protein for which we first took the PDB structure of 4B3E, used its fasta format and made change at the fifth position substituting T for A.There were 154 amino acids in each chain, as it is a complex it will exist in 2 chains i.e. A chain and B chain. After substitution the fasta format was used to search for templates in the SWISS-MODEL which can found in the website https://swissmodel.expasy.org and is a homology modeling software/tool. After the search was complete, 37 templates were found out of which 2 were having good GMQE and QMEAN scores, so we selected those templates to build our model (fig.2). The 2nd model build from the temple 2zkx.1.A was having z-score less than 1 (fig.3) and a z-

score of less than 1 represents a well predicted model as per the SWISS-MODEL guidelines.

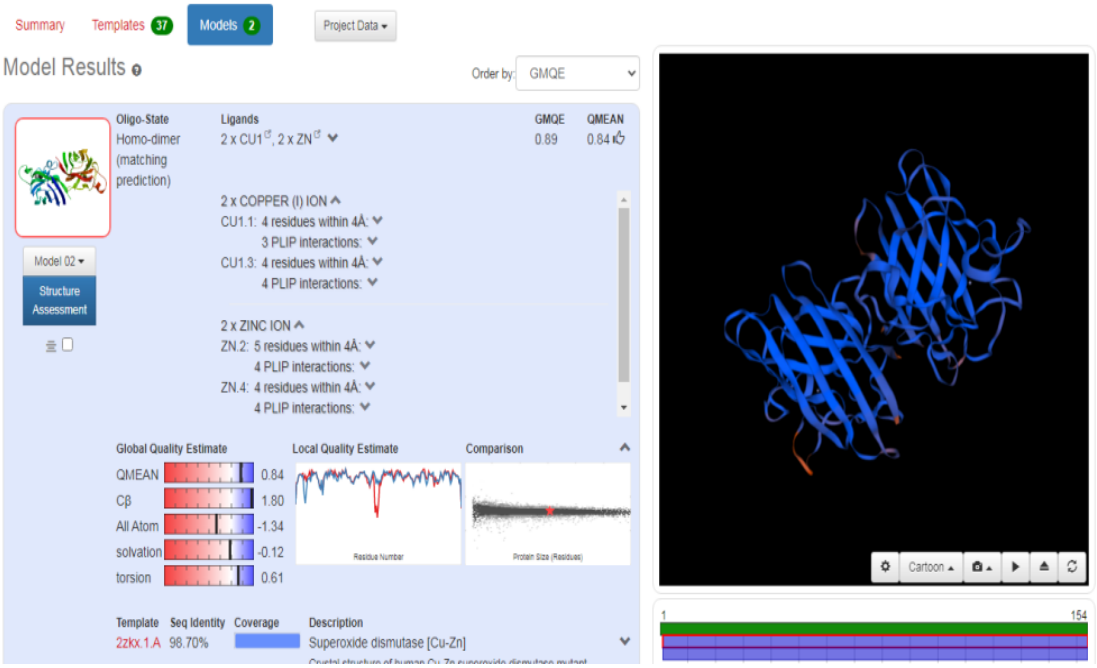

Fig.2. Model with GMQE and QMEAN scores

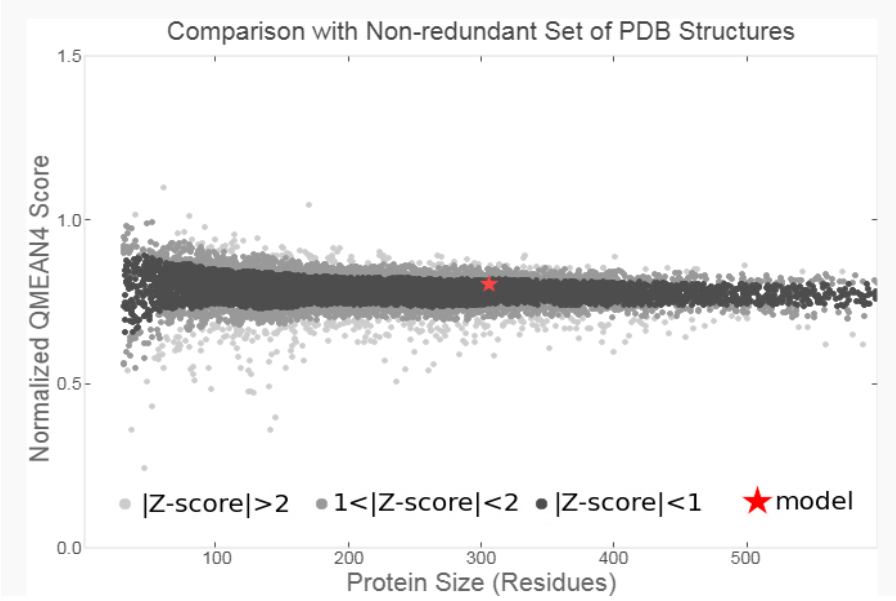

Fig.3. Z score of the new model

So, the 2nd model was selected as the mutant protein for our further studies. In order to distinguish mutant and wild-type protein we named the mutant protein as m4B3E (fig.4). Both the mutated protein and wild type protein simulation where carried out in two workstations simultaneously

2. After protein models were prepared, both molecular dynamics simulation were carried out using two work systems with similar configuration. The works systems were having Intel i5 processer, 16GB RAM with inbuilt GPU and LINUX operating system.

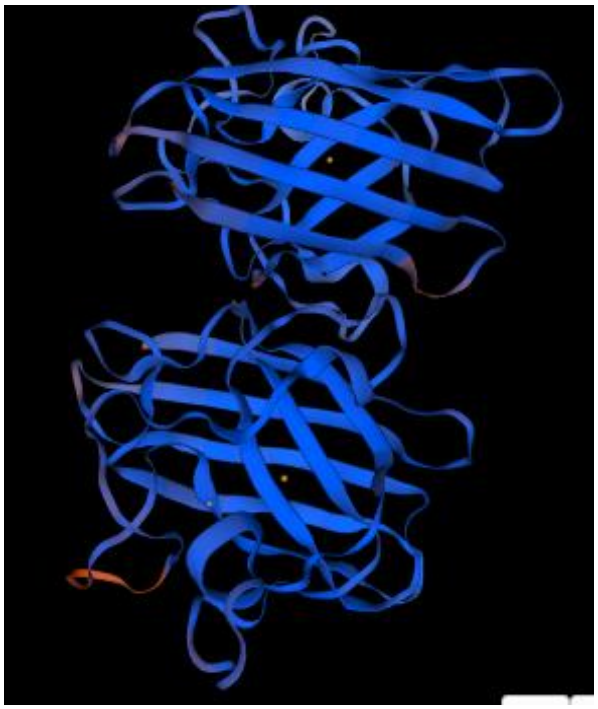

Fig.4. Mutant protein (m4B3E) model build using homology modeling

3. The first step of M D simulation starts with cleaning the raw data and removal of heteroatom from the .pdb file format followed by addition of water and selection of force field, there are about 15 force fields in GROMACS to choose from, we opted the 15th one which was the OPLS-AA/L all atom force field after which topology file was generated for the protein in both the workstations.

4. After the generation of topology we added dimension box in order to cover the whole protein inside it after which solvation was added to both the systems by GROMACS. For the mutant protein added solvation was 22166 and for the wild type it was 22434.

5. The next step is addition of ions to the solvated system, in case of mutated protein the system added 12 sodium ions where as is in the case of wild type protein it added 10 sodium ions. The ions were added in order balance the system.

6. After the creation of initial states, it was time for energy minimization so in order to carry out the energy minimization we used the command as per the MD tutorials. Energy minimization is done to find the minimum energy point at which the configuration of the system is most stable. For energy minimization it is not required that the system will run all the steps as per the command instead it will stop at the step with lowest energy coordinates. We Ran the energy minimization for 1000000 steps i.e. 2000 picoseconds but the system stop the process at 1531 steps for mutated protein and at 753 steps for wild-type protein. Further the energy minimizations were analysed using graphs to understand either the system's energy is minimized or not. We selected the potential plot (Energy against time plot) for both the protein systems and visualized the graph in GRACE software, full analysis of the combined graph can be seen in the results.

7. Once the energy of system was minimized it was time to equilibrate the system using ensembles, the two most common ensembles in use are NVT ensemble which has constant Number of particles, Volume, and Temperature also called as canonical ensemble and NPT ensemble which has constant Number of particles, Pressure, and Temperature also called as isothermal-isobaric ensemble. The NVT ensemble is used for stabilizing the system's temperature, so in order to do the same we used the NVT commands, ran it for 1 nanosecond and represented in the form of a temperature plot using software GRACE. Similarly NPT ensemble is used for stabilizing the pressure of the system for which we used the NPT commands, also run it for 1 nanosecond and represented it using a pressure plot, the analysis of which is given in the results.

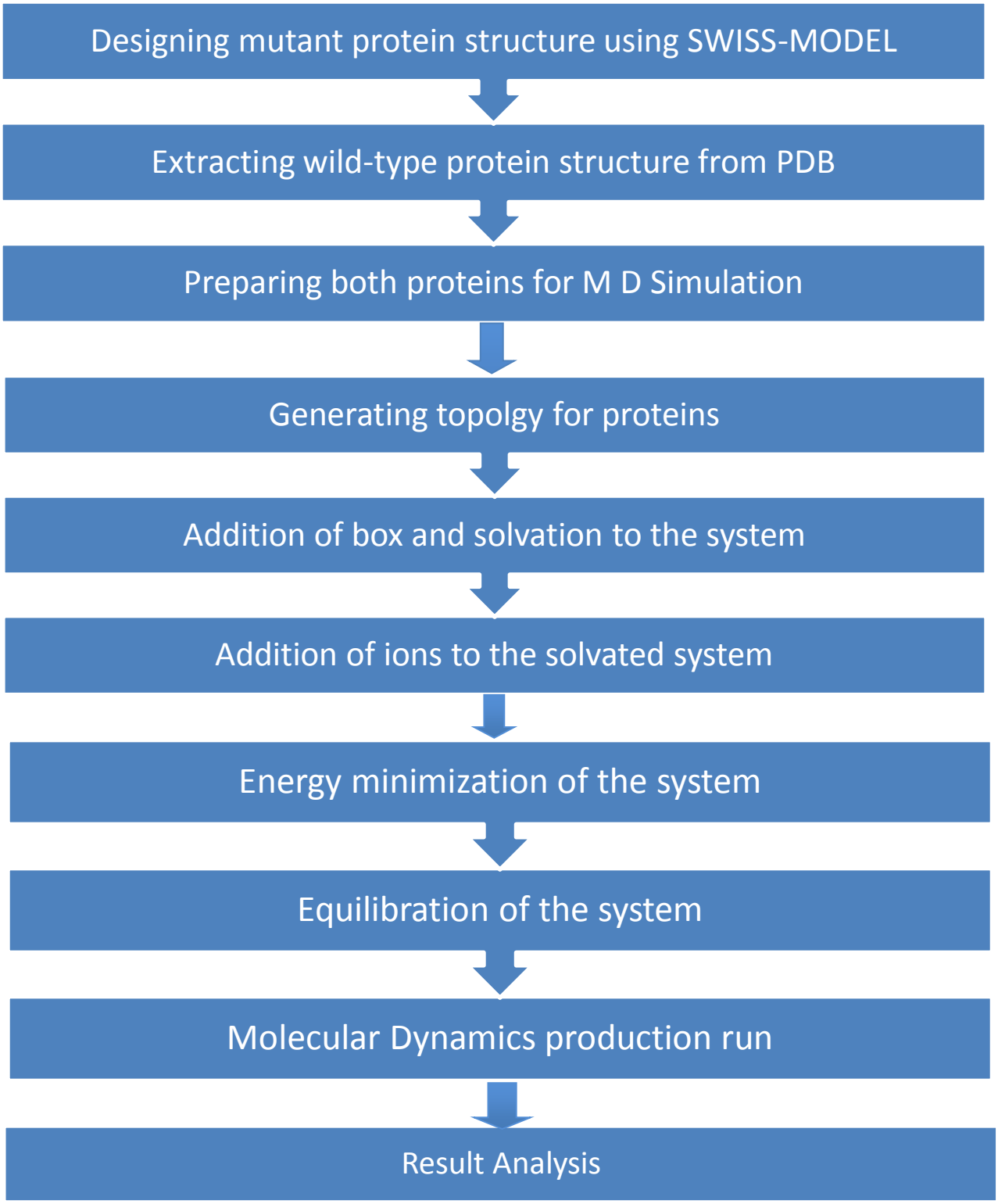

Fig..5. Methodology flowchart

8. All the above steps were used to stabilize the system and to prepare it for the main M D production run which is the last step of molecular dynamics simulation. To begin the process we provided the M D production Run commands and set the n-steps to 45,000,000 i.e. 90 nanoseconds. It took about 13 hours to complete the process in both the workstations. The results produced from M D Simulations were analysed using RMSD, RMSF and radius of gyration plots. The analyses of these graphs are provided in the result section of this paper. Further, in order to visualize the behavior of the simulated protein in 3 dimensions, we used the tool VMD (Visual Molecular Dynamics) refer to fig. 11, 12,13 and 14.

RESULTS

1. ENERGY MINIMIZATION ANALYSIS- . Energy minimization is done to find the minimum energy point at which the configuration of the system is most stable. Energy of the system is considered minimized as the graph show descending plot which means the energy is decreasing gradually. The energy minimization was carried out for 1000000 steps i.e. 2000 picoseconds but the system stop the process at 1531 steps for mutant protein and at 753 steps for wild-type protein which implies that at 1531th step the energy of the system was minimum for mutant protein and at 753th step the energy of the system was minimum for wild-type protein. Potential plot was generated for both the systems in GRACE software and the value of potential energy was -1.2531200e+06 for mutant protein and -1.2498100e+06 for wild type protein. The values of potential are evident that the energy of both systems is good to move forward with the experiment. A combined potential graph is shown in fig.5 in which the curve in black color represents potential energy of mutant protein and the curve in red color indicates potential energy of wild type protein.

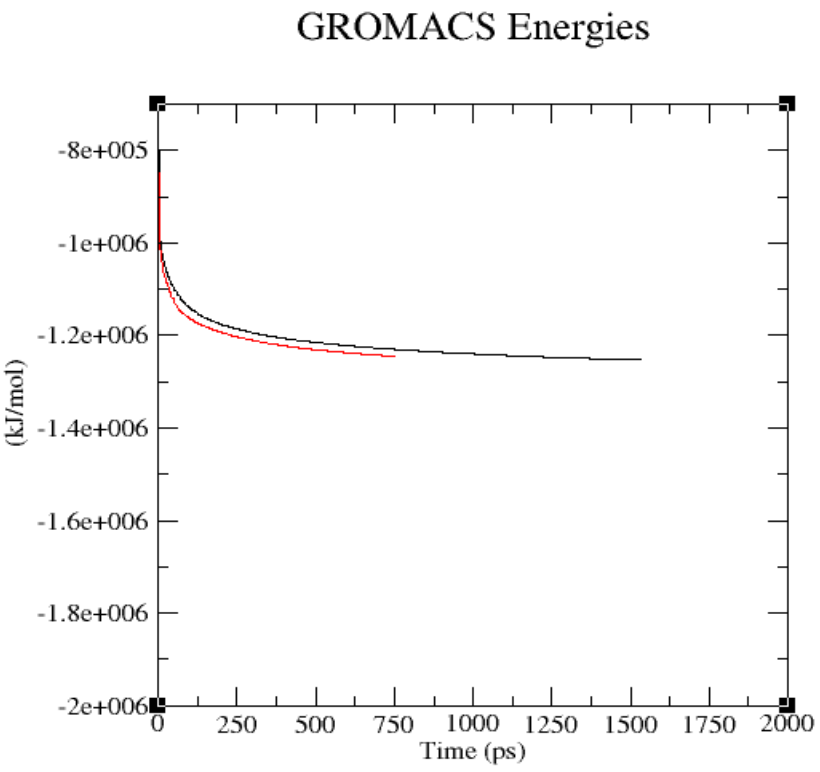

Fig.6. Combined potential graph of mutant (black) and wild type (red) protein

2. ANALYSIS OF SYSTEM EQUILIBRATION- The two most common ensembles used for system equilibration are NVT ensemble/canonical ensemble and NPT ensemble/isothermal-isobaric ensemble.

a) NVT ensemble/canonical ensemble- NVT stands for constant Number of particles, Volume, and Temperature. The NVT ensemble is used for stabilizing the temperature of the system. In this case we ran the NVT for 1ns in both the workstations and the temperature graph was plotted. In both the graphs the temperature fluctuation was between 294 and 305 which is very much considerable with this much simulation time. The temperature fluctuations of both the proteins were very close and stable. A combined graph of temperature is shown in Fig.6 in which the fluctuating black line indicates the temperature of mutant protein and red line indicates the temperature of wild type protein.

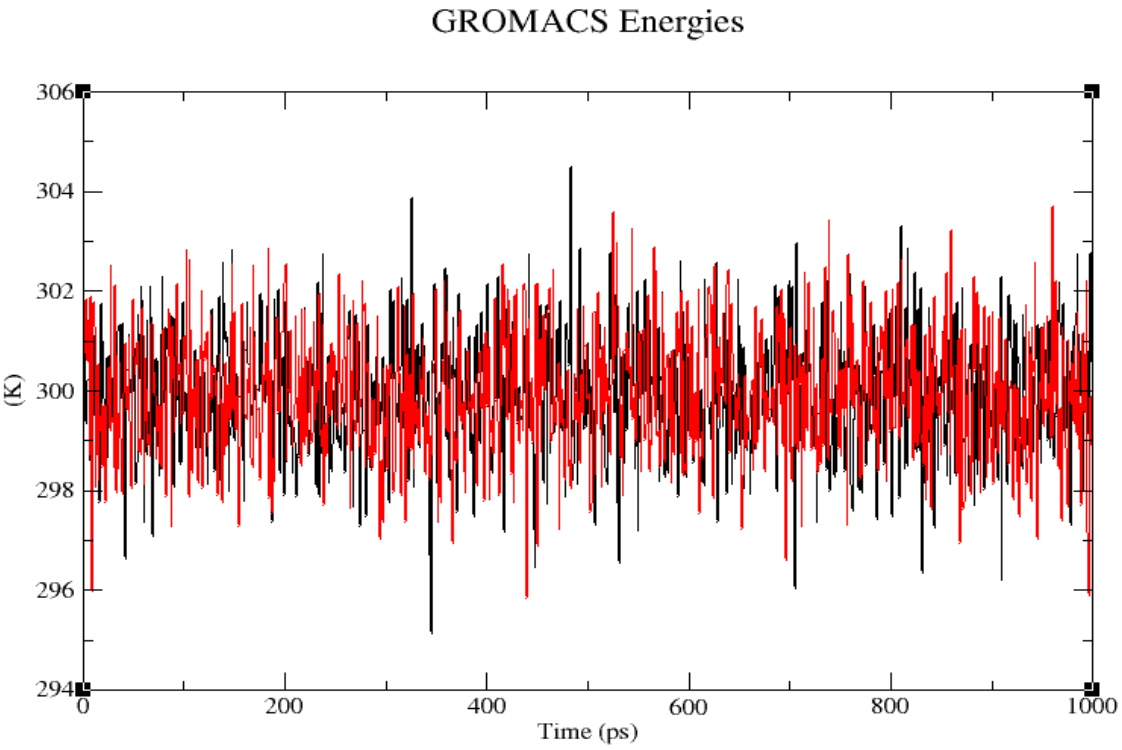

Fig.7. Combined temperature graph of mutant (black) and wild type (red) protein

b) NPT ensemble/ isothermal-isobaric ensemble- NPT stands for constant Number of particles, Pressure, and Temperature. It is used for stabilizing the system's pressure. In

this case we ran the NPT for 1ns in both the workstations and the pressure graph was plotted. In both the graphs the pressure fluctuation was between (–) 400 and (+) 400 which is very much considerable with this much simulation time. The pressure fluctuations of both the proteins were very close and stable. A combined graph of pressure is shown in Fig.7 in which the black line represents the pressure of mutant protein and red line represents the pressure of wild type protein.

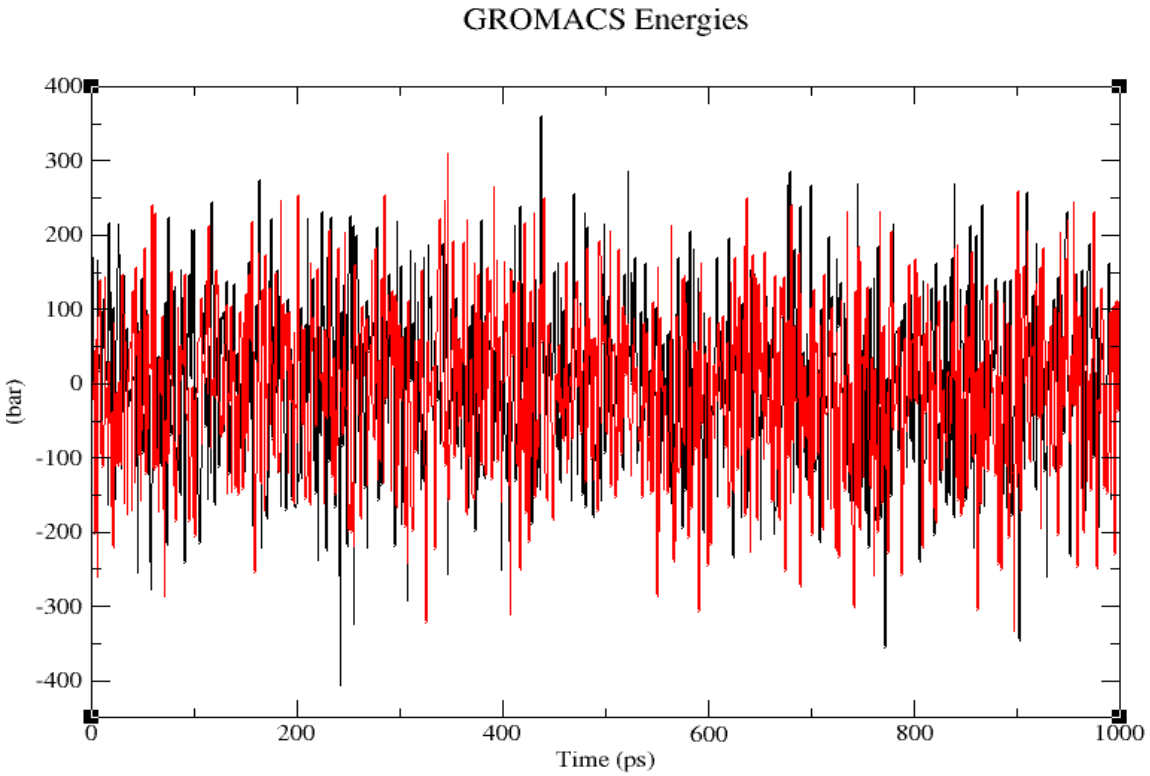

Fig.8. Combined pressure graph of mutant (black) and wild type (red) protein

3. RMSD ANALYSIS- RMSD is the abbreviation for Root Mean Square Deviation, which is used to measure the correctness of the model or to compare two conformations of a model at atomic level, in simple words it comprises value of average distance measured amongs the atoms of the protein model. RMSD can be calculated for protein backbone or for the whole protein, here the GROMACS has calculated the RMSD for protein backbone. A combined graph of RMSD is shown in Fig.8. in which the fluctuating black line indicates the mutant protein and red line indicates the wild type or experimentally verified protein. Considering the fact that the 4B3E protein's (wild type) structure is experimentally proven and fully stable as it was taken from PDB. So, the red line implies

the stable wild type protein RMSD values and while comparing it to the mutant protein RMSD values it can be observed that till 50ns the mutant structure was pretty much stable as the black line was overlapping red line (it shows the predicted mutant protein structure was right) but as the simulation was continued for long time the mutant protein structure became unstable which represents the fact the mutant protein is unstable for long run and is also expected to lose its function as it happened in the case of FALS.

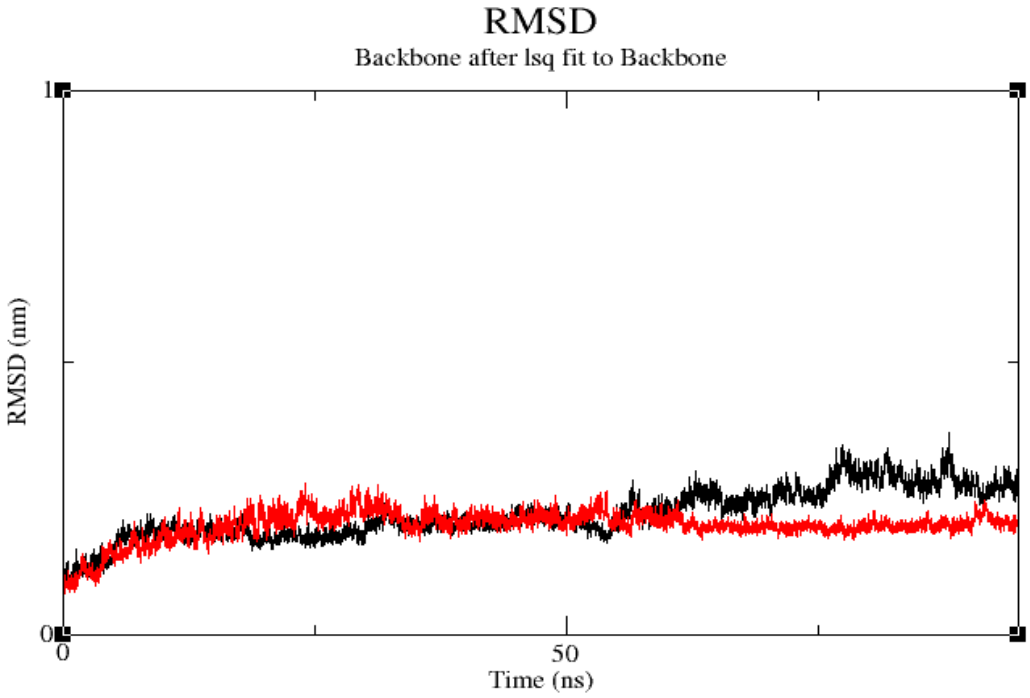

Fig.9. RMSD plot of mutant (black) and wild type (red) protein

4. RMSF ANALYSIS- RMSF is the abbreviation for Root Mean Square Fluctuations which can be expressed as a measure of deviation between position of particle and time averaged reference position. A RMSF graph was plotted using GRACE in both the workstations and later on combined for analysis. The combined RMSF plot is shown in Fig.9. which support the fact that mutant protein is unstable. From the graph we can observe that the black line is not overlapping the red line (experimentally verified 4B3E protein) this implies that the standard deviation of mutant protein (m4B3E) is not similar to the wild type (4B3E) protein and hence the stability is also not the same i.e. mutant protein is unstable.

5. RADIUS OF GYRATION ANALYSIS- Radius of gyration a measurement method used for measuring protein's compactness. A steady value of Rg (radius of gyration) represents a stably folded protein or vice-versa. For the graph shown in Fig.10. it is clearly observable that the black line (m4B3E) is less steady as compared to the

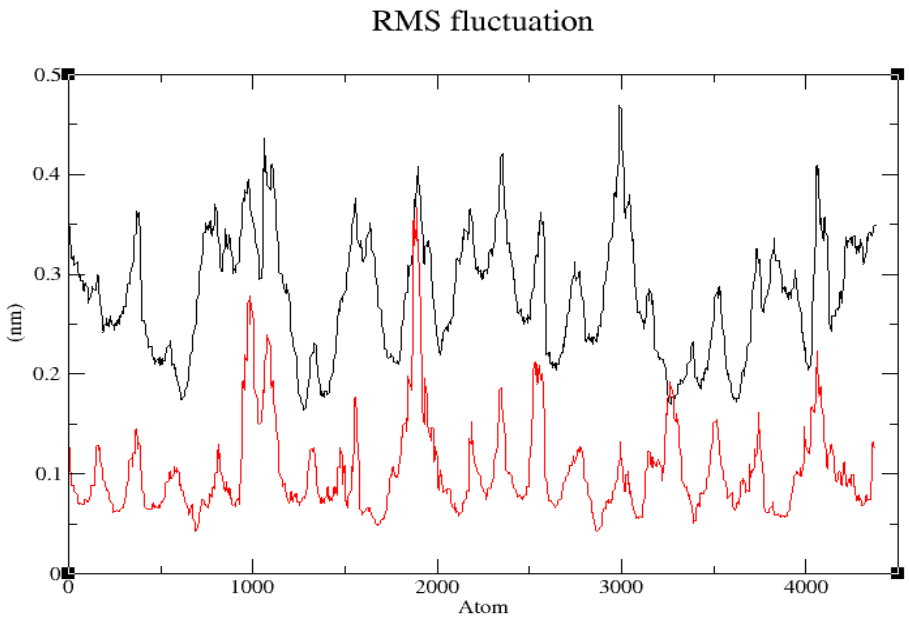

Fig.10. RMSF plot of mutant (black) and wild type (red) protein

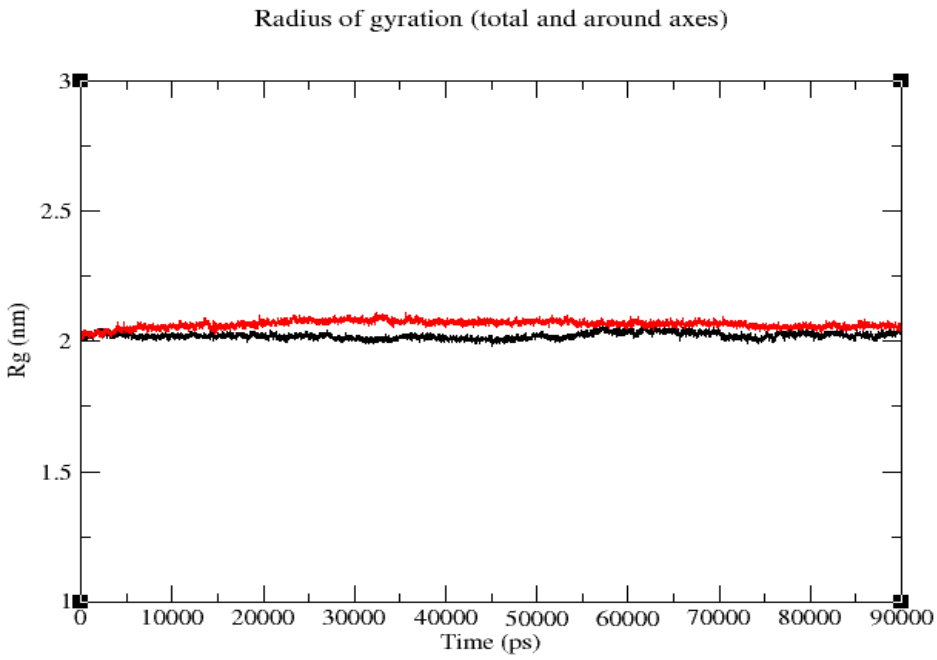

Fig.11. Radius of gyration plot of mutant (black) and wild type (red) protein

red line (4B3E) which implies that the m4B3E protein is not stably folded and hence, is losing its stability during the course of simulation. Similar to RMSD plot, in Rg plot the unsteadiness of the mutant protein became more relevant and observable after 50ns which means the protein became unstable after 50ns.

6. VISUALIZATION OF PROTEIN- The visualization was done using VMD (Visual Molecular Dynamics) tool. As discussed earlier the mutant protein and wild type protein both were stable till 50ns which can be observed in Fig.11 and 12. During further course of simulation i.e. after 50ns, the mutant protein became unstable and got unfolded/ refolded few times. The simulation visuals of both the proteins at time 75ns is shown in Fig.13 and 14.

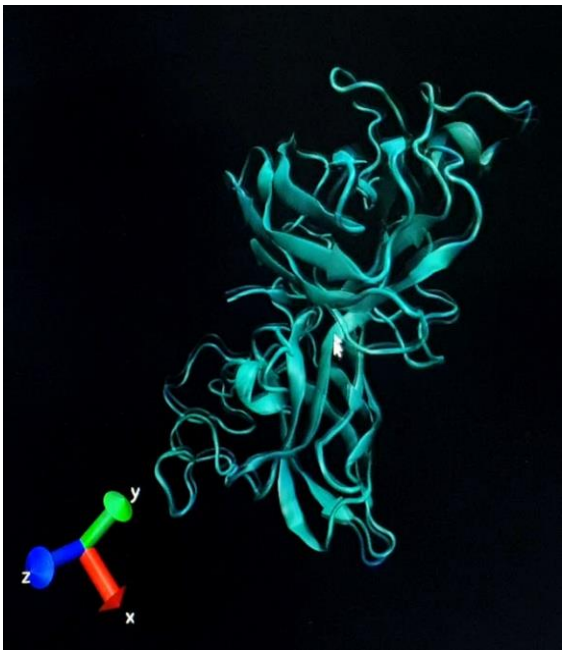

Fig.12. Wild type protein at time 45ns

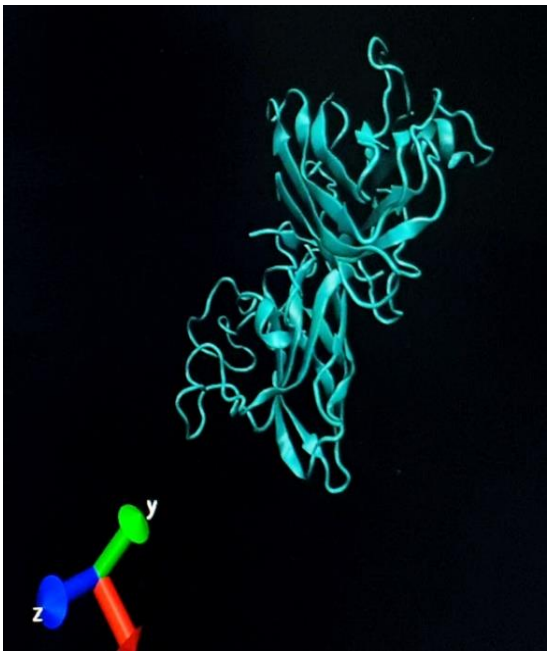

Fig.13. Mutant protein at time 45ns

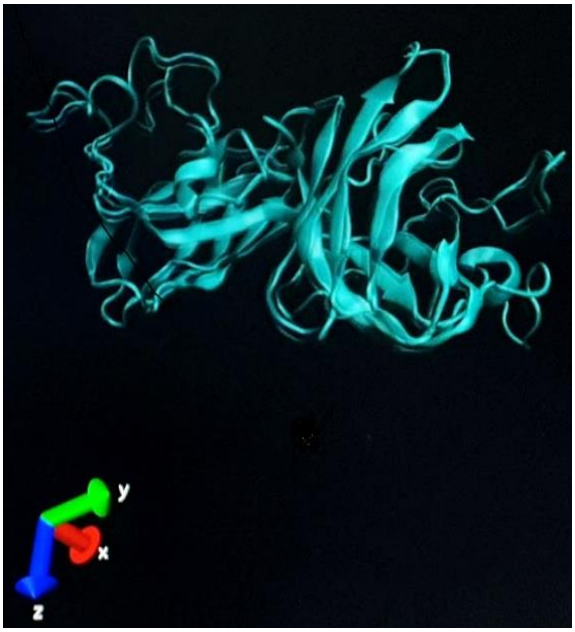

Fig.14. Wild type protein at time 75ns

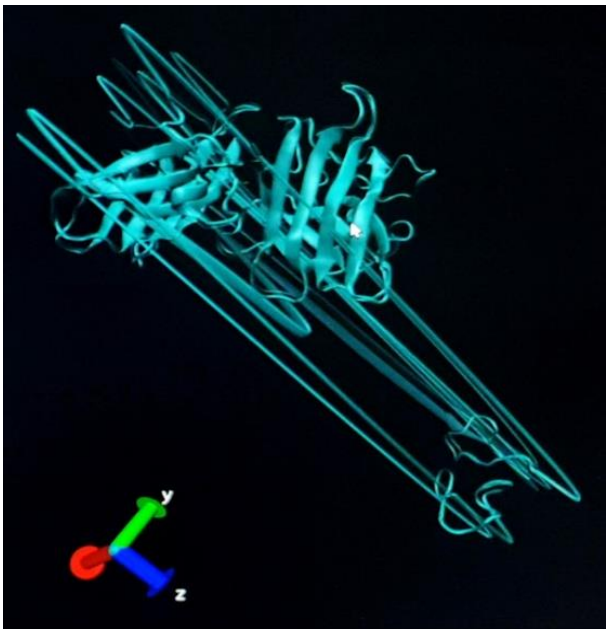

Fig.15. Mutant protein at time 75ns

DISCUSSION

We started with a published literature[4] which provided the information that a missense mutation of SOD1 gene at 5th position (substitution of T for A) was discovered in two patients of FALS belonging to one family. The published literature also concluded the possibility of involvement of this missense mutation of SOD1 gene in causing FALS. We wanted to support the statement by studying the behavior of the mutant protein in Molecular dynamics simulation which mimics the in vivo environmental conditions. As we predicted the mutant protein (m4B3E) using homology modeling software SWISS MODEL, the z-score suggested that the predicted is good and further to prove the stability of the predicted structure we carried out energy minimization and equilibration steps which also suggested that the model is stable. During the course of Molecular dynamics production run it was observed that the mutant protein structure was stable till 50nanoseconds which suggests the fact that there was nothing wrong in the predicted structure but as we moved further with our M D simulation it was observed that the after 50nanoseconds the structure became unstable till the final time i.e. 90nanoseconds which proved the fact that the mutant structure is actually unstable and may behave the same inside human body causing disorders as FALS. In the results section we have analysed each graph and came to a conclusion that the mutant structure is unstable in the long run, we also visualized the whole simulation process using VMD tool which also shows that the mutant protein structure became unstable after 50ns and also unfolded few times. This research article proves the fact that the missense mutation of SOD1 gene at 5th position in which A is substituted by T makes the superoxide dismutase enzyme unstable and malfunctioning which causes FALS.

FUTURE PROSPECTS

A lot is to be explored in terms of the FALS causes. The missense mutation discussed earlier i.e. missense mutation of SOD1 gene substituting Threonine (Thr) for Alanine (Ala) at the 5th position can be assessed further to bring more association of this mutation to other neuro-degenerative diseases and also to provide more evidences of its association to ALS and FALS in the wet lab experiments. In terms of dry lab experiment of the same, the mutant protein stability can be checked for further long time duration using Supercomputers in order to understand the behavior of the protein in long run and it will also give an idea about the progression of the disease according to the age of the patients, which can be very fruitful in real life scenario. As far as the M D Simulation is concerned, it's future is growing spontaneously and is expected to grow further more in the coming 25 years or so. MD simulation has its role in biological and chemical research sectors including drug designing, conformation identification, structure validation, model refinement, nanoscience etc. and is expected to become a vital part of these fields in the near future. Especially MD simulation is expected to bring massive growth in the drug discovery industry. As of now, the problems faced by the pharmaceutical industries in drug discovery are linked to the ever changing conformation of the proteins, target binging of the drug and validation of the drug in in-vivo conditions, all of these problems can be sorted out using MD simulation in the coming future which will further ensure its accuracy of mimicking in vivo conditions. MD simulation is a computer intensive procedure so, with the advancement of computer processer and GPUs, simulation will become more time efficient then what it is now. In the coming future more force fields are expected to be discovered in order to analyze more complex molecules. MD simulation is still in its initial phase as it's only about 60 years old, and there is a lot to explore further in terms of research and development in various sectors of science with a particular emphasis on biological and chemical studies.

REFERENCES

1. Rosen DR, Siddique T, Patterson D, Figlewicz DA, Sapp P, Hentati A, et al. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. ***Nature.*** 1993 Mar 4;**362**(6415):59–62. [PubMed] **[**Google Scholar**]**
2. Byrne S, Walsh C, Lynch C, et al. Rate of familial amyotrophic lateral sclerosis: a systematic review and meta-analysis. J Neurol Neurosurg Psychiatry 2011; 82:623.
3. Kevin Boylan. Familial ALS. Neurol Clin. 2015 Nov; 33(4): 807–830. Published online 2015 Sep 8. doi: 10.1016/j.ncl.2015.07.001
4. R Nakano , S Sato, T Inuzuka, K Sakimura, M Mishina, H Takahashi, F Ikuta, Y Honma, J Fujii, N Taniguchi, et al. A novel mutation in Cu/Zn superoxide dismutase gene in Japanese familial amyotrophic lateral sclerosis. Biochem Biophys Res Commun. 1994 Apr 29;200(2):695-703. doi: 10.1006/bbrc.1994.1506.
5. Kato S, Shaw P, Wood-Allum C, Leigh PN, Shaw CE: Amyotrophic lateral sclerosis. In Neurodegeneration: The molecular pathology of dementia and movement disorders Edited by: Dickson DW. ISN Neuropath press; 2003:350-371.
6. Lokesh C Wijesekera and P Nigel Leigh; Amyotrophic lateral sclerosis. Orphanet Journal of Rare Diseases. Orphanet Journal of Rare Diseases 2009, doi:10.1186/1750-1172-4-3
7. Rowland LP, Shneider NA: Amyotrophic lateral sclerosis. N Engl J Med 2001, 344:1688-1700.
8. Davenport RJ, Swingler RJ, Chancellor AM, Warlow CP: Avoiding false positive diagnoses of motor neuron disease: lessons from the Scottish Motor Neuron Disease Register. J Neurol Neurosurg Psychiatry 1996, 60:147-151.
9. Brooks BR: El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and the El Escorial "Clinical limits of amyotrophic lateral sclerosis" workshop contributors. J Neurol Sci 1994, 124(Suppl):96-107.
10. Brooks BR, Miller RG, Swash M, Munsat TL: El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyotroph Lateral Scler Other Motor Neuron Disord 2000, 1:293-299.

11. de Carvalho M, Dengler R, Eisen A, England JD, Kaji R, Kimura J, Mills K, Mitsumoto H, Nodera H, Shefner J, Swash M: Electrodiagnostic criteria for diagnosis of ALS. Clin Neurophysiol 2008, 119:497-503.
12. Worms PM: The epidemiology of motor neuron diseases: a review of recent studies. J Neurol Sci 2001, 191:3-9.
13. Sejvar JJ, Holman RC, Bresee JS, Kochanek KD, Schonberger LB: Amyotrophic lateral sclerosis mortality in the United States, 1979–2001. Neuroepidemiology 2005, 25:144-152.
14. Norris F, Shepherd R, Denys E, U K, Mukai E, Elias L, Holden D, Norris H: Onset, natural history and outcome in idiopathic adult motor neuron disease. J Neurol Sci 1993, 118:48-55.
15. Tyler HR, Shefner J: Amyotrophic lateral sclerosis. Handb Clin Neurol 1991:169-215.
16. Aran FA: Recherches sur une maladie non encore décrite du systemé musculaire (atrophie musculaire progressive). Arch Gen Med 1850, 14:5-35. 172–214
17. Duchenne de Boulogne GB: Recherches électro-physiologiques et thérapeutiques. Comp Rend Seances Acad Sci 1851, 32:506.
18. Cruveilhier J: Sur le paralysie musculaire, progressive, atrophique. Bull Acad Med 1853, 18:490-501. 546–583
19. Goetz CG: Amyotrophic lateral sclerosis: Early contributions of Jean-Martin Charcot. Muscle & Nerve 2000, 23:336-343.
20. Rowland LP: How Amyotrophic Lateral Sclerosis Got Its Name: The Clinical-Pathologic Genius of Jean-Martin Charcot. Arch Neurol 2001, 58:512-515.
21. de Carvalho M, Matias T, Coelho F, Evangelista T, Pinto A, Luis ML: Motor neuron disease presenting with respiratory failure. J Neurol Sci 1996, 139(Suppl):117-122.
22. Chen R, Grand'Maison F, Strong MJ, Ramsay DA, Bolton CF: Motor neuron disease presenting as acute respiratory failure: a clinical and pathological study. J Neurol Neurosurg Psychiatry 1996, 60:455-458.
23. Rosen DR, Siddique T, Patterson D, Figlewicz DA, Sapp P, Hentati A, Donaldson D, Goto J, O'Regan JP, Deng HX, et al.: Mutations in Cu/ Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature 1993, 362:59-62

24. Pamela J, Shaw PGI, Falkous Gavin, Mantle David: Oxidative damage to protein in sporadic motor neuron disease spinal cord. Annals of Neurology 1995, 38:691-695.
25. Ferrante RJ, Browne SE, Shinobu LA, Bowling AC, Baik MJ, MacGarvey U, Kowall NW, Brown RH Jr, Beal MF: Evidence of increased oxidative damage in both sporadic and familial amyotrophic lateral sclerosis. J Neurochem 1997, 69:2064-2074.
26. R Glenn Smith YKH, Mark P Mattson, Stanley H Appel: Presence of 4-hydroxynonenal in cerebrospinal fluid of patients with sporadic amyotrophic lateral sclerosis. Annals of Neurology 1998, 44:696-699.
27. Hideo Tohgi TA, Kinya Yamazaki, Takahiko Murata, Eri Ishizaki, Chiaki Isobe: Remarkable increase in cerebrospinal fluid 3-nitrotyrosine in patients with sporadic amyotrophic lateral sclerosis. Annals of Neurology 1999, 46:129-131.
28. Figlewicz DA, Krizus A, Martinoli MG, Meininger V, Dib M, Rouleau GA, Julien J-P: Variants of the heavy neurofilament subunit are associated with the development of amyotrophic lateral sclerosis. Hum Mol Genet 1994, 3:1757-1761.
29. Tomkins J, Usher P, Slade JY, Ince PG, Curtis A, Bushby K, Shaw PJ: Novel insertion in the KSP region of the neurofilament heavy gene in amyotrophic lateral sclerosis (ALS). Neuroreport 1998, 9:3967-3970.
30. Shaw PJ: Molecular and cellular pathways of neurodegeneration in motor neurone disease. J Neurol Neurosurg Psychiatry 2005, 76:1046-1057.
31. Cozzolino M, Ferri A, Carri MT: Amyotrophic lateral sclerosis: from current developments in the laboratory to clinical implications. Antioxid Redox Signal 2008, 10:405-443.
32. Desport JC, Preux PM, Magy L, Boirie Y, Vallat JM, Beaufrere B, Couratier P: Factors correlated with hypermetabolism in patients with amyotrophic lateral sclerosis. Am J Clin Nutr 2001, 74:328-334.
33. Kasarskis EJ, Berryman S, Vanderleest JG, Schneider AR, McClain CJ: Nutritional status of patients with amyotrophic lateral sclerosis: relation to the proximity of death. Am J Clin Nutr 1996, 63:130-137.
34. Leigh PN, Abrahams S, Al-Chalabi A, Ampong MA, Goldstein LH, Johnson J, Lyall R, Moxham J, Mustfa N, Rio A, et al.: The management of motor neurone disease. J Neurol Neurosurg Psychiatry 2003, 74(Suppl 4):iv32-iv47.

35. Heffernan C, Jenkinson C, Holmes T, Macleod H, Kinnear W, Oliver D, Leigh N, Ampong MA: Management of respiration in MND/ ALS patients: an evidence based review. Amyotroph Lateral Scler 2006, 7:5-15.
36. Miller RG, Rosenberg JA, Gelinas DF, Mitsumoto H, Newman D, Sufit R, Borasio GD, Bradley WG, Bromberg MB, Brooks BR, et al.: Practice parameter: The care of the patient with amyotrophic lateral sclerosis (an evidence-based review): Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 1999, 52:1311.
37. Lechtzin N, Wiener CM, Shade DM, Clawson L, Diette GB: Spirometry in the supine position improves the detection of diaphragmatic weakness in patients with amyotrophic lateral sclerosis. Chest 2002, 121:436-442.
38. Gruis KL, Brown DL, Schoennemann A, Zebarah VA, Feldman EL: Predictors of noninvasive ventilation tolerance in patients with amyotrophic lateral sclerosis. Muscle Nerve 2005, 32:808-811.
39. Schmidt EP, Drachman DB, Wiener CM, Clawson L, Kimball R, Lechtzin N: Pulmonary predictors of survival in amyotrophic lateral sclerosis: use in clinical trial design. Muscle Nerve 2006, 33:127-132.
40. Rahman, A. Correlations in the motion of atoms in liquid argon. *Phys Rev*,1964; 136, A405.
41. Barker, J.A. and Watts, R.O. Structure of water; A Monte Carlo calculation. *Chem Phys Lett*, 1969; 3(3), pp.144-145.
42. McCammon, J.A., Gelin, B.R., Karplus, M. Dynamics of folded proteins. *Nature*, 1977; 267,585-590.
43. Parrinello, M. and Rahman, A. Crystal structure and pair potentials: A molecular dynamics study. *Phys Rev Lett*, 1980; 45, 1196.
44. Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *J Chem Phys*, 1980; 72, 2384-2393.
45. Day, R., Paschek, D. and Garcia, A.E. Microsecond simulations of the folding/unfolding thermodynamics of the Trp-cage miniprotein. *Proteins*, 2010; 78, 1889-1899.
46. Lopes, P.E., Guvench, O., MacKerell, A.D. Jr. Current status of protein force fields for molecular dynamics simulations. *Methods Mol Biol*, 2015; 1215, 47-71.

47. Jorgensen, W.L., Tirado-Rives, J. The OPLS Force Field for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J Am Chem Soc*, 1988; 110, 1657–1666.
48. Van Duin, .C.T., Dasgupta, S., Lorant, F., Goddard, W.A. ReaxFF: A Reactive Force Field for Hydrocarbons. *J Phys Chem A*, 2001; 105 (41), 9396–9409.
49. Ponder, J.W., Wu, C., Ren, P., Pande, V.S., Chodera, J.D., Schnieders, M.J., Haque, I., Mobley, D.L., Lambrecht, D.S., DiStasio, R.A. Jr, Head-Gordon, M., Clark, G.N., Johnson, M.E., Head-Gordon, T. Current status of the AMOEBA polarizable force field. *J Phys Chem B*, 2010; 114(8), 2549-2564.
50. Oostenbrink, C., Villa, A., Mark, A.E., Van Gunsteren, W.F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem*, 2004; 25(13), 1656–1676.
51. MacKerell, A.D., Banavali, N. and Foloppe, N.Development and current status of the CHARMM force field for nucleic acids. Biopolymers, 2000; 56, 257-265.
52. Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T.E., Laughton, C.A. and Orozco, M. Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophys J*, 2007; 92, 3817-3829.
53. Selvaraj, C., Sakkiah, S., Tong, W., Hong, H., Molecular dynamics simulations and applications in computational toxicology and nanotoxicology, *Food and Chemical Toxicology* (2017), doi: 10.1016/j.fct.2017.08.028.
54. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput*. 2008;4(3):435–447.
55. Weiner, P.K. and Kollman, P.A. AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. *J Comput Chem*, 2, 1981; 287-303.
56. Nelson, M.T., Humphrey, W., Gursoy, A., et al. NAMD: a parallel, object oriented molecular dynamics program. *Int J Supercomput Appl High Perform Comput*. 1996;10(4):251–268.

57. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comp Chem*, 1983; 4 (2), 187–217.
58. Kevin J. Bowers, Edmond Chow, Huafeng Xu, Ron O. Dror, Michael P. Eastwood, Brent A. Gregersen, John L. Klepeis, István Kolossváry, Mark A. Moraes, Federico D. Sacerdoti, John K. Salmon, Yibing Shan, and David E. Shaw, "Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters," *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)*, Tampa, Florida, November 11–17, 2006.
59. Joshua A Rackers , Zhi Wang , Chao Lu , Marie L Laury , Louis Lagardère , Michael J Schnieders , Jean-Philip Piquemal , Pengyu Ren , Jay W Ponder , Tinker 8: Software Tools for Molecular Design. J Chem Theory Comput. 2018 Oct 9;14(10):52735289. doi: 10.1021/acs.jctc.8b00529.
60. Paul Martin, Steve Parker, Dino Spagnoli, Arnaud Marmier, Dean Sayle, et al.. Application of Molecular Dynamics DL_POLY Code to Interfaces of Inorganic Materials. Molecular Simulation, Taylor & Francis, 2007, 32 (12-13), pp.1079-1093. ff10.1080/08927020601013817.
61. Pitman, M. R., & Menz, R. I. Methods for Protein Homology Modelling. Applied Mycology and Biotechnology, 2006; 37–59. doi:10.1016/s1874-5334(06)80005-5.
62. Adam Hospital, Josep Ramon Goñi, Modesto Orozco, Josep L Gelpí. Molecular dynamics simulations: advances and applications. *Dovepress journal.2015*
63. Harvey MJ, Giupponi G, De Fabritiis G. ACEMD: accelerating biomolecular dynamics in the microsecond time scale. *J Chem Theory Comput*. 2009;5(6):1632–1639.
64. Fan, H. and Mark, A.E. Refinement of homology-based protein structures by molecular dynamics simulation techniques. Protein Sci, 2004; 13: 211-220.
65. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. How fast-folding proteins fold. *Science*. 2011;334(6055):517–520.
66. M. Wieder, U. Perricone, T. Seidel, S. Boresch and T.Langer, *Monatshefte fur Chemie*, 2016, **147**, 553–563.
67. M. Gorska-Ponikowska, U. Perricone, A. Kuban-Jankowska, G. Lo Bosco and G. Barone, *J. Cell. Physiol.*, DOI:10.1002/jcp.25888.

68. M. Wieder, U. Perricone, S. Boresch, T. Seidel and T.Langer, *Biochem. Biophys. Res. Commun.*, 2016, **470**, 685–689.
69. G. Deganutti and S. Moro, *Molecules*, 2017, **22**, 818.
70. A. Cuzzolin, M. Sturlese, G. Deganutti, V. Salmaso, D. Sabbadin, A. Ciancetta and S. Moro, *J. Chem. Inf.Model.*, 2016, **56**, 687–705.
71. G. Deganutti, A. Welihinda and S. Moro, *ChemMedChem*, 2017, **12**, 1319–1326.
72. S. Decherchi, A. Berteotti, G. Bottegoni, W. Rocchia and A. Cavalli, *Nat. Commun.*, 2015, **6**, 6155.
73. A. Magistrato, J. Sgrignani, R. Krause and A. Cavalli, *J. Phys. Chem. Lett.*, 2017, **8**, 2036–2042.
74. R. Gaspari, C. Rechlin, A. Heine, G. Bottegoni, W. Rocchia, D. Schwarz, J. Bomke, H.-D. Gerber, G. Klebe and A. Cavalli, *J. Med. Chem.*, 2016, **59**, 4245–4256.
75. U. Perricone, M. R. Gulotta, J. Lombino, B. Parrino, S. Cascioferro, P. Diana, G. Cirrincione and A. Padova, An overview of recent Molecular Dynamics applications as medicinal chemistry tool for undruggable sites challenge. *Med. Chem. Commun*., 2018, DOI: 10.1039/C8MD00166A.
76. Raval A, Piana S, Eastwood MP, Dror RO, Shaw DE. Refinement of protein structure homology models via long, all-atom molecular dynamics simulations. *Proteins*. 2012;80(8):2071–2079.
77. Bryn Fenwick R, Esteban-Martin S, Salvatella X. Understanding biomolecular motion, recognition, and allostery by use of conformational ensembles. *Eur Biophys J*. 2011;40(12):1339–1355.
78. Bowman GR, Huang X, Pande VS. Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods*. 2009; 49(2):197–201.
79. Kirchmair, J., Williamson, M.J., Tyzack, J.D., Tan, L., Bond, P.J., Bender, A. and Glen, R.C. Computational prediction of metabolism: sites, products, SAR, P450 enzyme dynamics, and mechanisms. *J Chem Inf Model*, 2012; 52, 617-648.
80. Egan, W.J., Zlokarnik, G. and Grootenhuis, P.D. In silico prediction of drug safety: despite progress there is abundant room for improvement. Drug Discov Today, 2004; 1, 381-387.

81. Bock JE, Gavenonis J, Kritzer JA. Getting in shape: controlling peptide bioactivity and bioavailability using conformational constraints. ACS Chem Biol 2013;8: 488–99. https://doi.org/10.1021/cb300515u.
82. Merten C, Li F, Bravo-Rodriguez K, Sanchez-Garcia E, Xu Y, Sander W. Solventinduced conformational changes in cyclic peptides: a vibrational circular dichroism study. Phys Chem Chem Phys 2014;16:5627–33. https://doi.org/10.1039/C3CP55018D.
83. Henninot A, Collins JC, Nuss JM. The current state of Peptide drug discovery: back to the future? J Med Chem 2017. https://doi.org/10.1021/acs.jmedchem.7b00318 acs.jmedchem.7b00318.
84. Ho BK, Dill KA. Folding very short Peptides usingmolecular dynamics. PLoS Comput Biol 2006;2:e27. https://doi.org/10.1371/journal.pcbi.0020027.
85. Voelz VA, Shell MS, Dill KA. Predicting Peptide structures in native proteins from physical simulations of fragments. PLoS Comput Biol 2009;5:e1000281. https://doi.org/10.1371/journal.pcbi.1000281.
86. Damas JM, Filipe LCS, Campos SRR, Lousa D, Victor BL, Baptista AM, et al. Predicting the thermodynamics and kinetics of Helix Formation in a Cyclic Peptide model. J Chem Theory Comput 2013;9:5148–57. https://doi.org/10.1021/ct400529k.
87. Razavi AM, Wuest WM, Voelz VA. Computational screening and selection of cyclic peptide hairpin mimetics by molecular simulation and kinetic network models. J Chem Inf Model 2014;54:1425–32. https://doi.org/10.1021/ci500102y.
88. Wang J,Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. J Comput Chem 2004;25:1157–74. https://doi.org/10.1002/jcc.20035.
89. Butterfoss GL, Yoo B, Jaworski JN, Chorny I, Dill KA, Zuckermann RN, et al. De novo structure prediction and experimental characterization of folded peptoid oligomers. Proc Natl Acad Sci 2012;109:14320–5. https://doi.org/10.1073/pnas.1209945109.
90. Ogorzalek TL, Hura GL, Belsom A, Burnett KH, Kryshtafovych A, Tainer JA, et al. Small angle X-ray scattering and cross-linking for data assisted protein structure prediction in CASP 12 with prospects for improved accuracy. Proteins Struct Funct Bioinforma 2018;86:202–14. https://doi.org/10.1002/prot.25452.

91. Wolynes PG, Friedrichs MS. Toward Protein tertiary structure recognition by means of associative memory Hamiltonians. Science 1989;246:371–3. https://doi.org/10.1126/science.246.4928.371.
92. Friedrichs MS, Goldstein RA, Wolynes PG. Generalized protein tertiary structure recognition using associative memory hamiltonians. J Mol Biol 1991;222:1013–34. https://doi.org/10.1016/0022-2836(91)90591-S.