Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 9679 - 9685
Received 25 April 2021; Accepted 08 May 2021.

In-silico Analysis and Molecular Dynamics Simulations of Lysozyme by GROMACS 2020.2

Anil Panwar[1], Ashok Kumar[1*]
[1,1*]Centrefor Systems Biology and Bioinformatics, Panjab University,
Chandigarh-160014, India. Email id: ashokbiotech@gmail.com

Abstract
Since last few years, number of publicationsdescribing accurate and reliable all-atom molecular dynamics simulations got significantly hiked. Availability of faster computers, development of high precision and fast methods for calculating the long-range electrostatic interactions made it possible. With the advances in hardware and software, it has been demonstrated that molecular dynamics simulation is not only capable of mimicking the behavior of macro molecules, but also can also be used for bio molecular structure prediction and drug discovery.Proteins are highly dynamic structures and their dynamism contributes toward ligand binding properties.*In-vivo* analysis of protein dynamism is very complex, expensive and tedious task. Therefore scientific community has lots of hope with *in-silio* methods. Present study uses MD simulations to explore relation between Sequence conservation and dynamism of amino acids of Lysozyme.
Three Dimensional Structures of Five Lysozyme were downloaded from RCSB PDB. Amino Acid Sequences of download PDB structures were aligned using CLUSTAL O. Newton's equation of motion was solved by considering all atoms simulation method. GROMACS 2020.2 package was used to perform MD simulations and all atom OPLS force field was used. GROMACS module Pdb2gmx was used to generate the topology of protein. Simple point charge water model [SPC216] was used to solvate the protein. Protein was solvated to maintain the equilibrium. The equilibrated system was then minimized at maximum force of 1000.0 KJ/mol/nm by using 50,000 steps. The solvated and energy minimized systems were than equilibrated for 100ps under NVT and NPT ensemble processes. All the bonds were constrained by LINCS algorithm. Finally 1ns molecular dynamics simulation was run to observe the stability of proteins. The Root Mean Standard Fluctuation was calculated and their 2D graphs were than plotted with xmgrance software. Comparative Root Mean Square Fluctuation (RMSF) values were found to be least for the segment whichis largest reserve segment in the sequences under course of study. RMSF for Lysozyme class shows a reciprocal relation with sequence conservation length.

Keywords
Lysozyme, Molecular Dynamics, Sequence Conservation, Molecular Simulations.

1. Introduction

Proteins are highly dynamic structures. *In-vivo* analysis of protein dynamism is very complex, expensive and tedious task. Therefore scientific community has lots of hope with *in-silio* methods of protein dynamism. Molecular Dynamics mimic what atoms do in real life, assuming a given potential energy functions. Some discoveries like discovery regarding the molecule myoglobin, which could have been made only by using MD (1). Molecular dynamics (MD) simulations predict how every atom in a protein or other molecular system will move over time based on a general model of the physics governing interatomic interactions (2).The energy function allows us to calculate the force experienced by any atom given the positions of the other atoms and the Newton's laws tell us how those forces will affect the motions of the atoms. Alder and Wainwright in the 1950's study the interactions of hard spheres (3). They were the first to introduce the molecular

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 9679 - 9685
Received 25 April 2021; Accepted 08 May 2021.

dynamics. The behavior of simple liquids revealed by their study (4). The first simulation was carried out by Rahman in 1964 using a realistic potential for liquid argon (5). The simulation of liquid water was performed by Rahman and Stillinger in 1974 (6). That was considered as first molecular dynamics simulation of a realistic system. Simulations carried out by McCammon*et al.* in 1977 on bovine pancreatic trypsin inhibitor (BPTI) was considered as first protein simulations (7).Structure prediction has been one of the most ancient problems addressed in structural bioinformatics. MD, including the longest simulations performed, has been extensively used for *ab initio* protein structure prediction (8-11).

2.0 Material and Methodology

2.1 Data Collection, Compilation and Multiple Sequence Alignment

Tertiary structures (3D images) of Five Lysozyme were obtained from Research Collaboratory for Structural Bioinformatics Protein data bank (RCSB PDB) (12). This PDB database contains 169963 Biological Macromolecular Structures, mostly of them (150477) are resolved by X-Ray crystallography. Out of 169963 entries, 49498 entries belong to Homo sapiens. Data compilation was done by editing tertiary structure files. Crystal water and heteroatoms were stripped out. Protein Sequences in FASTA format were also obtained from the PDB database. Multiple sequences Alignment (MSA) was done by Clustal O (13).

2.2 Molecular Dynamics Simulations

MD Simulations was done by GROMACS 2020.2 (14). PDB fileswere verified that all the necessary atoms should present. Topology fileswere made; topology contains all the information necessary to define the molecule within a simulation. This information includes non-bonded parameters (atom types and charges) as well as bonded parameters (bonds, angles, and dihedrals). The force field information wasalso written to the topology file. A unit cell was constructed and filled by water using solvate module. Ions were added according to charge present on the protein. The solvated, electro neutral systemswent for molecular dynamic simulations. Before the dynamics begins, the systemswere checked for steric clashes or inappropriate geometry. The structureswere relaxed through a process called energy minimization (EM). After ensuring that the systemswere at an energy minimum state, real dynamics began.

Protein Equilibration was conducted in two phases. The first phase was conducted under an NVT ensemble (constant Number of particles, Volume, and Temperature). This ensemble is also referred to as "isothermal-isochoric". The timeframe for such a procedure is dependent upon the contents of the system, but in NVT, the temperature of the system should reach a plateau at the desired value. Temperature was set to a maximum of 310 K. we conducted a 100-ps NVT equilibration. NVT equilibration stabilized the temperature of the system. Equilibration of pressure is conducted under an NPT ensemble, wherein the Number of particles, Pressure, and Temperature are all constant. We conducted a 100-ps NPT equilibration. After completion of the two equilibration phases, the systemswere well-equilibrated at the desired temperature and pressure and ready to run MD for data collection. We run a 1-ns MD simulation with a time step of 100; means for every 10^{-11} second, a trajectory image was recorded.The Root Mean Standard Fluctuation was calculated and their 2D graphs were than plotted with xmgrance software (15).

3.0 Results

3.1 Retrieval and Analysis of 3D structures

Three Dimensional Structures of Five Lysozyme Described in Table 1 were downloaded from RCSB PDB. Human Lysozymewas found to be longest with 130 amino acids, while with 129 amino acids

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 9679 - 9685
Received 25 April 2021; Accepted 08 May 2021.

rest four sequences were of similar length. Hetero atoms and water were removed using "grep" command of linux.

Table: 1. Lysozyme used under the course of study.

S. No.	PDB ID	Source Organism	Protein Sequence length
1.	1AKI	*Gallus gallus*	129
2.	2Z2F	*Bostaurus*	129
3.	1LMN	*Oncorhynchusmykiss*	129
4.	4D9Z	*Gallus gallus*	129
5.	1REX	*Homo sapiens*	130

3.2 Multiple Sequence alignment

Multiple Sequence alignment (MSA) of all five Lysozyme Amino Acid sequences was done using CLUSTAL O. All five sequences were found to possess great sequence conservation (Fig 1).

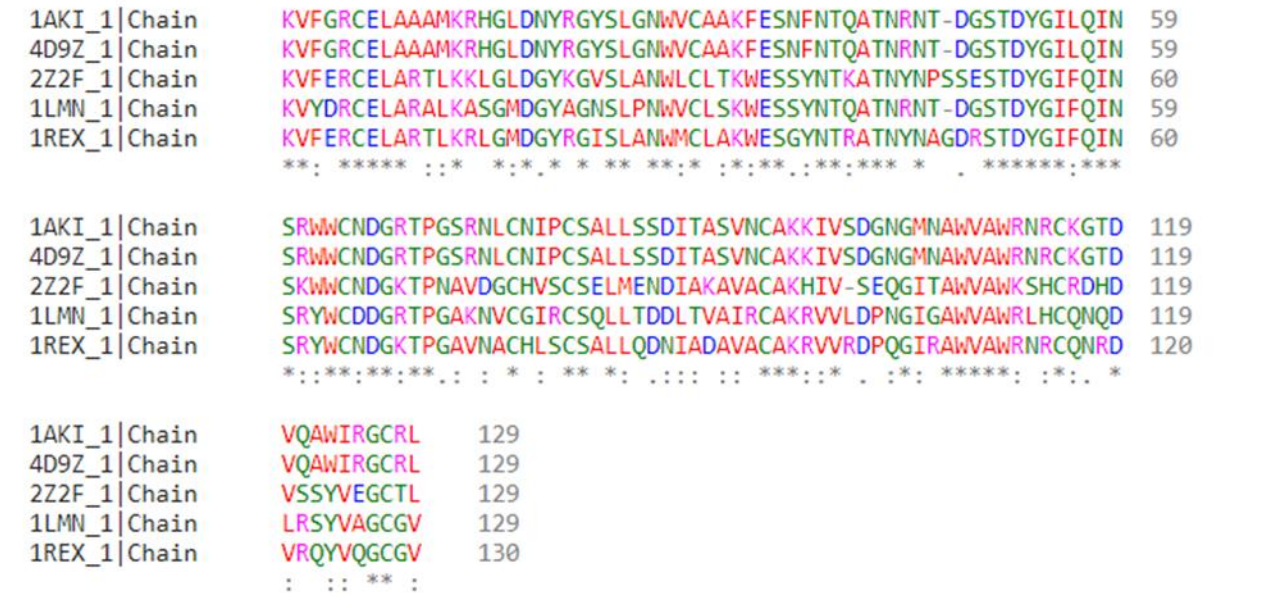


Fig: 1. Multiple Sequence Alignment of Lysozyme sequences.

3.3 Molecular Dynamics Simulations

OPLS-AA/L all atom force field was used to generate Topology file.A unit cell (box) is constructed and protein is placed at the centre of the box, and it places at least 1.0 nm from the box edge. Ions were added according to charge present on the protein.The electro neutral structures were relaxed through a process called energy minimization (Fig3). After ensuring that the system is at an energy minimum state, real dynamics began. Protein Equilibration was done under NVT and NPT. Temperature was set to a maximum of 310 K. we conducted a 100-ps NVT equilibration (Fig5). NVT equilibration stabilized the temperature of the system. Prior to data collection, we stabilized the pressure of the system. Equilibration of pressure is conducted under NPT (Fig4). We conducted a 100-ps NPT equilibration. After completion of the two equilibration phases, we run MD for data collection. We run a 1-ns MD simulation witha time step of 100; means for every 10^{-11}second, a trajectory image was recorded. Root mean square fluctuations (RMSF) of residues during the course of study were recorded (Fig5).

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 9679 - 9685
Received 25 April 2021; Accepted 08 May 2021.

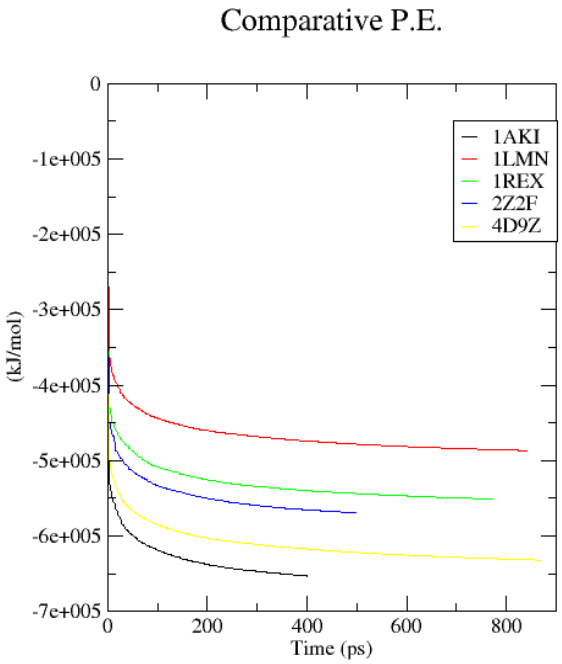


Fig: 2. Comparative P.E. Steps of Systems

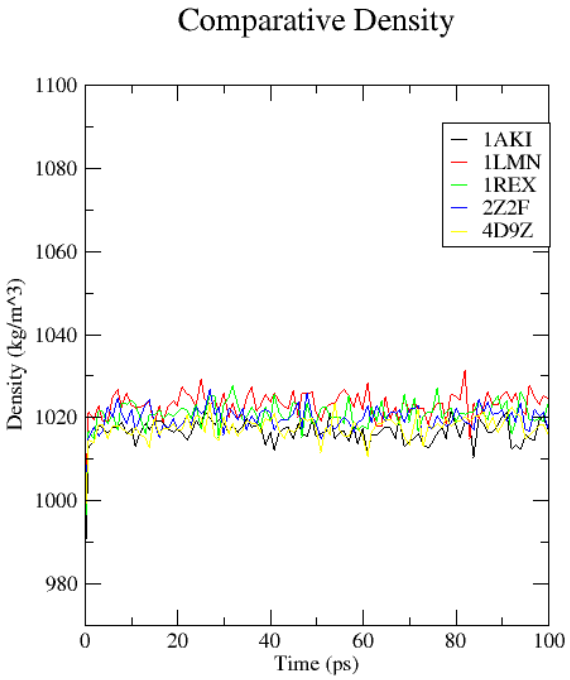


Fig: 3. Comparative Density of Systems

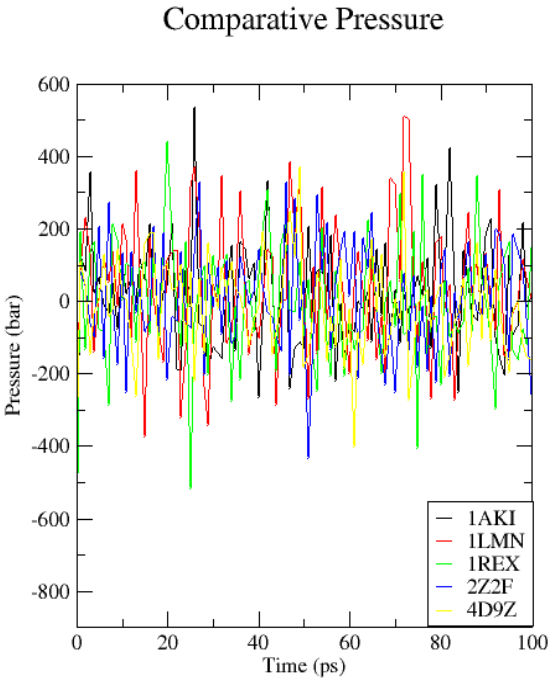


Fig: 4. NPT graph of Systems

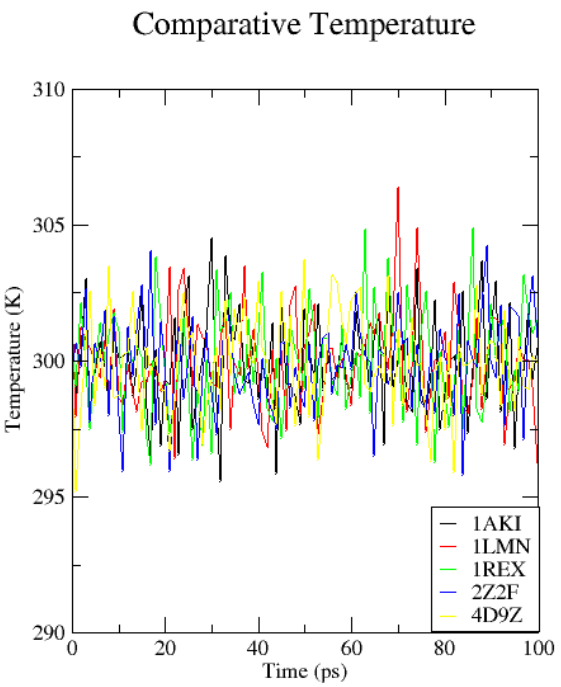


Fig: 5. NVT graph of Systems

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 9679 - 9685
Received 25 April 2021; Accepted 08 May 2021.

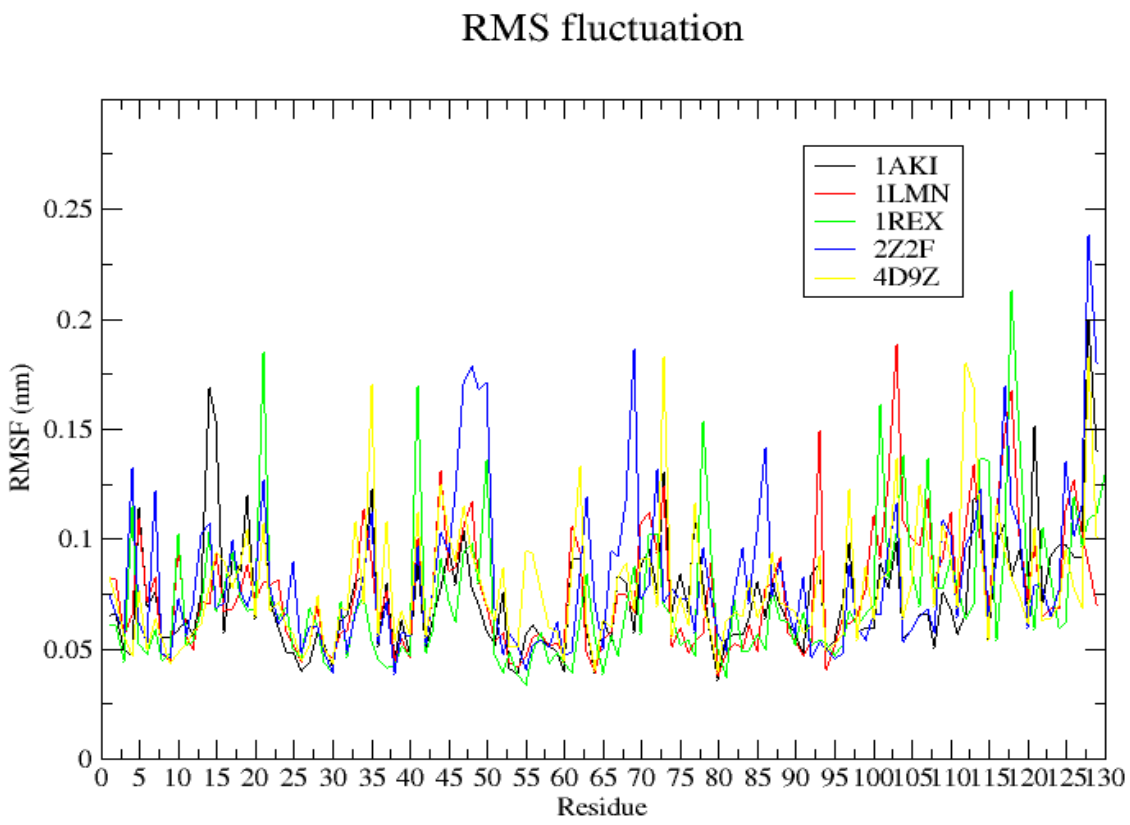
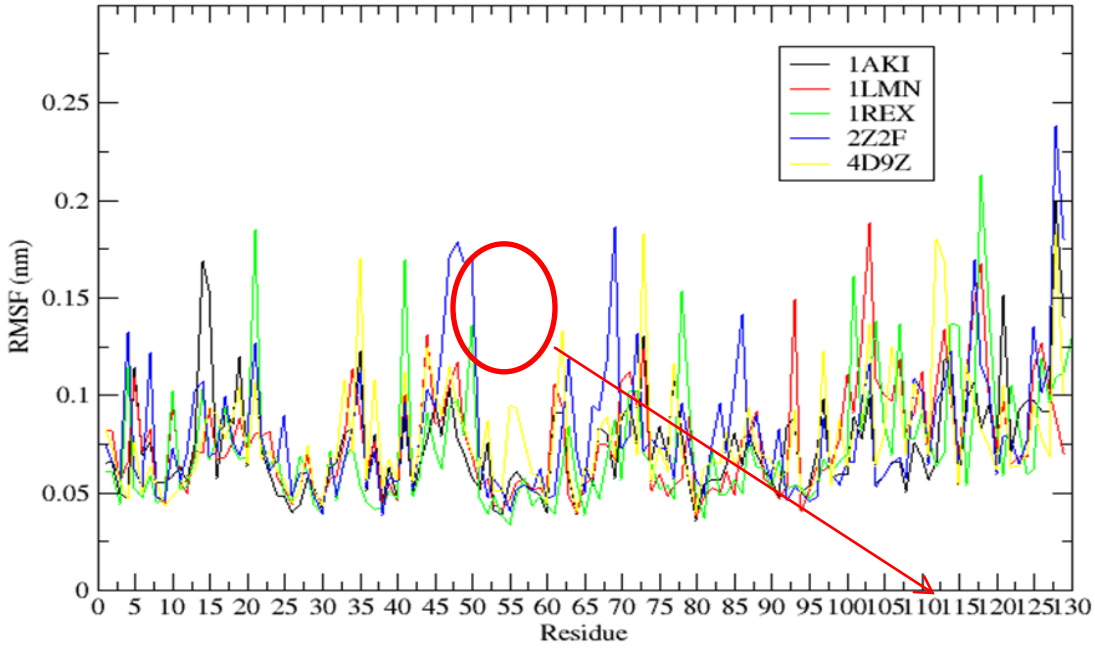


Fig: 6. ComparativeRMSF graph of Systems

4.0 Discussion

The longest stretch of sequences which exhibits conservation was found at places 51-61 where 10 amino acids out of 11 were found reserved (Fig1). When we observe the RMSF at these places, we found that these places have least RMSF values (Fig7). The highest peaks in RMSF plot, relates to those regions which possess least or no sequence conservation (Fig7). The least Dynamic and most sequence conserve part denotes strand (Fig8).



Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 9679 - 9685
Received 25 April 2021; Accepted 08 May 2021.

```
1AKI_1|Chain      KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNT-DGSTDYGILQIN  59
4D9Z_1|Chain      KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNT-DGSTDYGILQIN  59
2Z2F_1|Chain      KVFERCELARTLKKLGLDGYKGVSLANWLCLTKWESSYNTKATNYNPSSESTDYGIFQIN  60
1LMN_1|Chain      KVYDRCELARALKASGMDGYAGNSLPNWVCLSKWESSYNTQATNRNT-DGSTDYGIFQIN  59
1REX_1|Chain      KVFERCELARTLKRLGMDGYRGISLANWMCLAKWESGYNTRATNYNAGDRSTDYGIFQIN  60
                  **: ***** ::*  *:*.* * ** **:* :*:**.:**:*** *  . ******:***

1AKI_1|Chain      SRWWCNDGRTPGSRNLCNIPCSALLSSDITASVNCAKKIVSDGNGMNAWVAWRNRCKGTD  119
4D9Z_1|Chain      SRWWCNDGRTPGSRNLCNIPCSALLSSDITASVNCAKKIVSDGNGMNAWVAWRNRCKGTD  119
2Z2F_1|Chain      SKWWCNDGKTPNAVDGCHVSCSELMENDIAKAVACAKHIV-SEQGITAWVAWKSHCRDHD  119
1LMN_1|Chain      SRYWCDDGRTPGAKNVCGIRCSQLLTDDLTVAIRCAKRVVLDPNGIGAWVAWRLHCQNQD  119
1REX_1|Chain      SRYWCNDGKTPGAVNACHLSCSALLQDNIADAVACAKRVVRDPQGIRAWVAWRNRCQNRD  120
                  *::**:**:**.: : * : ** *: .::: :: ***::* . :*: *****: :*:. *

1AKI_1|Chain      VQAWIRGCRL  129
4D9Z_1|Chain      VQAWIRGCRL  129
2Z2F_1|Chain      VSSYVEGCTL  129
1LMN_1|Chain      LRSYVAGCGV  129
1REX_1|Chain      VRQYVQGCGV  130
                  :  :: ** :
```

Fig: 7. RMSF relating with Sequence conservation

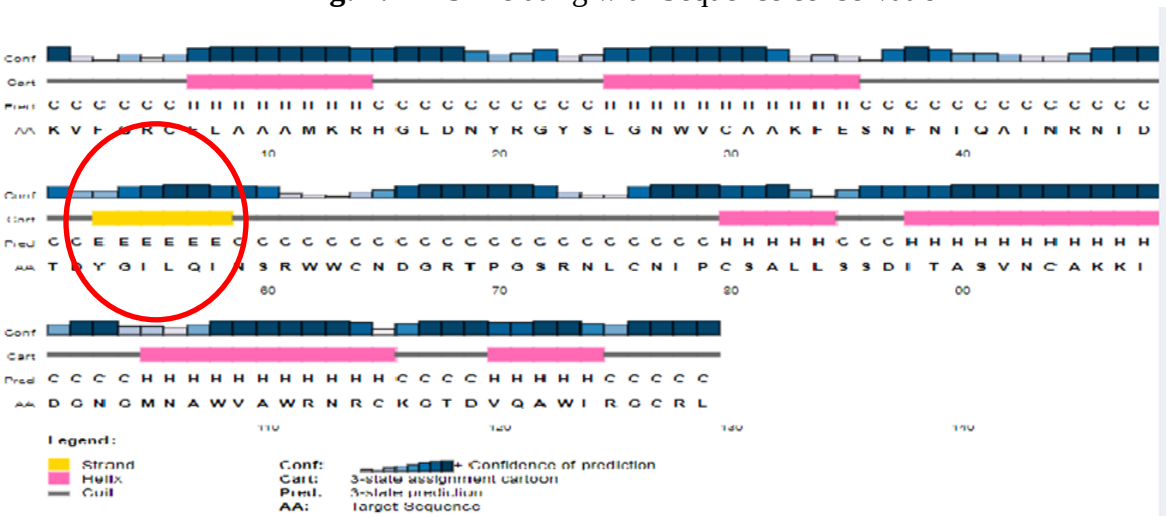


Fig: 8. The least Dynamic and most sequence conserve part denotes strand.

5.0 Conclusion

The aim of this study was to findrelation between amino acids sequence, secondary structure and dynamic properties of residues of Lysozyme class. Study shows a rational approach based on molecular simulations can successfully be used to study the dynamic nature of proteins. In this study we found that the places in the Lysozyme protein sequences which possess highest sequence conservation shows the least dynamic properties.It is concluded that part of enzymes sequences having highest sequence conservation are least dynamic in nature. The least dynamic and most reserve part in Lysozyme codes for strand in secondary structure.

6.0 Acknowledgement

Authors express their heartfelt thanks to the Centre for Systems Biology and Bioinformatics, Panjab University, Chandigarh for providing necessary infrastructure facility in collection, analysis and interpretation of data. We would also like to acknowledge the support provided by online tools, servers, databases and software for successful accomplishment of our work.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 6, 2021, Pages. 9679 - 9685
Received 25 April 2021; Accepted 08 May 2021.

References

1. Karplus, M. and McCammon, J.A., 1986. The dynamics of proteins. Scientific American, 254(4), pp.42-51.
2. Karplus, M., and McCammon, J.A. (2002). Molecular dynamics simulations ofbiomolecules. Nat. Struct. Biol. 9, 646–652.
3. Alder, B.J. and Wainwright, T.E., 1957. Phase transition for a hard sphere system. The Journal of chemical physics, 27(5), pp.1208-1209.
4. Alder, B.J. and Wainwright, T.E., 1959. Studies in molecular dynamics .I. general Method. The Journal of chemical physics, 31, 459, (https://doi.org/10.1063/1.1730376).
5. Rahman, A., 1964. Correlations in the motion of atoms in liquid argon. Physical Review, 136(2A), p.A405.
6. Stillinger, F.H. and Rahman, A., 1974. Improved simulation of liquid water by molecular dynamics. The Journal of Chemical Physics, 60(4), pp.1545-1557.
7. McCammon, J.A., Gelin, B.R. and Karplus, M., 1977. Dynamics of folded proteins. nature, 267.
8. Dorn, M., e Silva, M.B., Buriol, L.S. and Lamb, L.C., 2014. Three-dimensional protein structure prediction: Methods and computational strategies. Computational biology and chemistry, 53, pp.251-276.
9. Lindorff-Larsen, K., Piana, S., Dror, R.O. and Shaw, D.E., 2011. How fast-folding proteins fold. Science, 334(6055), pp.517-520.
10. Piana, S., Lindorff-Larsen, K. and Shaw, D.E., 2013. Atomic-level description of ubiquitin folding. Proceedings of the National Academy of Sciences, 110(15), pp.5915-5920.
11. Bonneau, R. and Baker, D., 2001. Ab initio protein structure prediction: progress and prospects. Annual review of biophysics and biomolecular structure, 30(1), pp.173-189.
12. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ...& Bourne, P. E. , 2000. The protein data bank. Nucleic acids research, 28(1), 235-242.
13. Sievers F, Higgins DG, 2002. Clustal Omega. Current Protocols in Bioinfor-matics. 48. John Wiley & Sons, Inc. pp. 3.13.1–16
14. Berendsen HJ, van der Spoel D, van Drunen R, 1995. GROMACS: a message-passing parallel molecular dynamics implementation. Computer physics communi-cations 91(1-3):43-56
15. Turner PJ. XMGRACE, Version 5.1.19, 2005.Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton