Advances in Bioresearch
Adv. Biores., Vol 12 (4) July 2021: 58-63
©2021 Society of Education, India
Print ISSN 0976-4585; Online ISSN 2277-1573
Journal's URL:http://www.soeagra.com/abr.html
CODEN: ABRDC3
DOI: 10.15515/abr.0976-4585.12.4.5863

Advances in Bioresearch

ORIGINAL ARTICLE

Insight into Relation between Sequence Conservation and Dynamic Properties of Amino Acids of Trypsin

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

ABSTRACT
Molecular Dynamics (MD) is a really useful tool in the hands of the modern scientist of computational expert. Through microscopic molecular dynamics simulations, macroscopic properties of a system are explored. Molecular Dynamics mimic what atoms do in real life. 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 a lots of hope with insilio methods. Present study uses MD simulations to explore relation between Sequence conservation and dynamism of amino acids of Trypsin. Three Dimensional Structures of Five Trypsin 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 Trypsin class of enzymes shows a reciprocal relation with sequence conservation length.
Keywords: *Molecular Dynamics; Trypsin; Sequence Conservation, Molecular Simulations*

Received 19.03.2021 Revised 21.04.2021 Accepted 11.05.2021

How to cite this article:
A Panwar, A Kumar. Insight into Relation between Sequence Conservation and Dynamic Properties of Amino Acids of Trypsin. Adv. Biores. Vol 12 [4] July 2021. 58-63

INTRODUCTION

Enzymes are catalyst which without being used regulates the rate of chemical reactions that takes place within the living organisms. Digestion takes place with the help of digestive juices. Digestive juices like gastric juice, pancreatic juice, intestinal and bile juice are enriched with many enzymes. Digestive enzymes fall under the class of Hydrolases, one among the six classes in which enzymes are being classified. One of them is trypsin which helps in digestion of proteins. 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.The energy function allows us to calculate the force experienced by any atom given the positions of theother atoms and the Newton's laws tell us how those forces will affect the motions of the atoms. Some discoveries like discovery regarding the molecule myoglobin, which could have been made only by using MD. Alder and Wainwright in the 1950's study the interactions of hard spheres. They were the first to introduce the molecular dynamics. The behavior of simple liquids revealed by their study. The first simulation was carried out by Rahman in 1964 using a realistic potential for liquid argon. The simulation of liquid water was performed by Rahman and Stillinger in 1974. 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. Long MD simulations extensively used for ab initio protein structure prediction

MATERIAL AND METHODS

Data Collection, Compilation and Multiple Sequence Alignment

Tertiary structures (3D images) of Five Trypsins were obtained from Research Collaboratory for Structural Bioinformatics Protein data bank (RCSB PDB). This PDB database contains 169436 Biological Macromolecular Structures, mostly of them (150077) are resolved by X-Ray crystallography. Out of 169436 entries, 49307 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.

Molecular Dynamics Simulations

MD Simulations was done by GROMACS 2020.2. PDB files were verified that all the necessary atoms should present. Topology files were 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 was also 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 systems went for molecular dynamic simulations. Before the dynamics begins, the systems were checked for steric clashes or inappropriate geometry. The structures were relaxed through a process called energy minimization (EM). After ensuring that the systems were 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 systems were 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.

RESULTS

Retrieval and Analysis of 3D structures

Three Dimensional Structures of Five Trypsins Described in Table 1 were downloaded from RCSB PDB. Human trypsin was found to be longest with 224 amino acids, while with 220 amino acids *Gadusmorhua* trypsin found to be smallest. Hetero atoms and water were removed using “grep” command of linux.

Table: 1. Trypsin Enzymes used under the course of study.

S. No.	PDB ID	Source Organism	Protein Sequence length
1.	1DPO	*Rattusrattus*	223
2.	1FN6	*Sus scrofa*	223
3.	1H4W	*Homo sapiens*	224
4.	2EEK	*Gadusmorhua*	220
5.	2ZPS	*Oncorhyn-chusketa*	222

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).

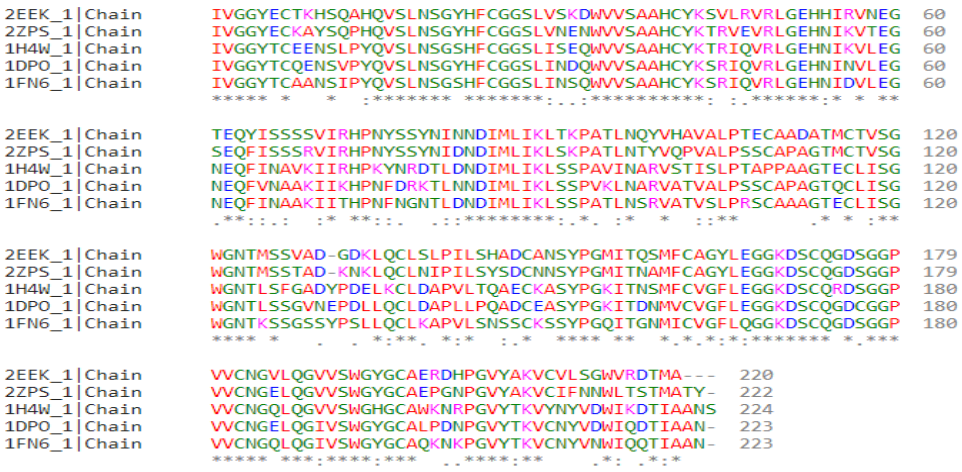

Fig1. Multiple Sequence Alignment of enzyme sequences.

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 (Table2).

Table2. Neutralization and Potential Energy Minimization steps of Subject proteins.

S.no.	PDB ID	Protein is Neutralised by	P.E. Steps
1.	1DPO	7 Na+ ions	974
2.	1FN6	5 Cl^- ions	931
3.	1H4W	1 Na+ ion	855
4.	2EEK	2 Na+ ions	1027
5.	2ZPS	1 Na+ ion	668

The electro neutral structures were relaxed through a process called energy minimization (Fig3). The P.E. Steps taken by each system is given in Table2. 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 with a 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).

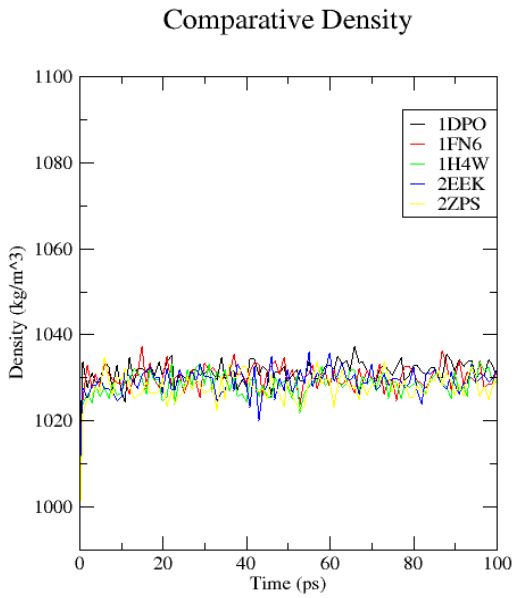

Fig2. Comparative Density of Systems

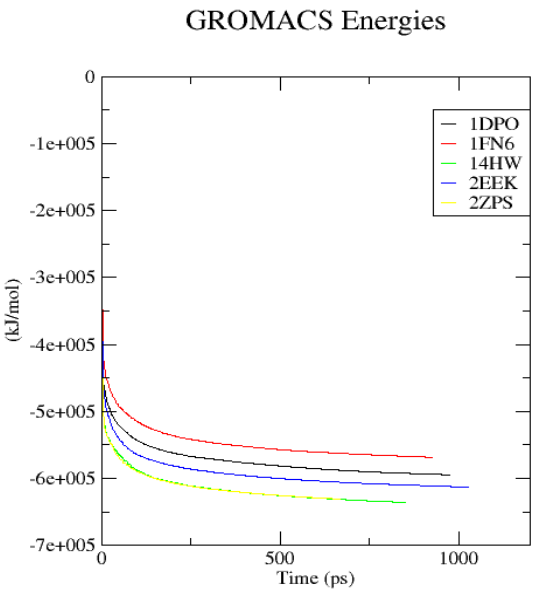

Fig3.P.E. Steps of Systems

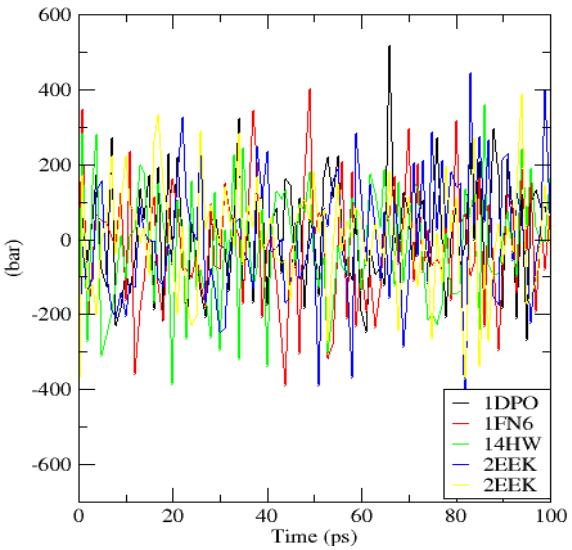

Fig4. NPT graph of Systems

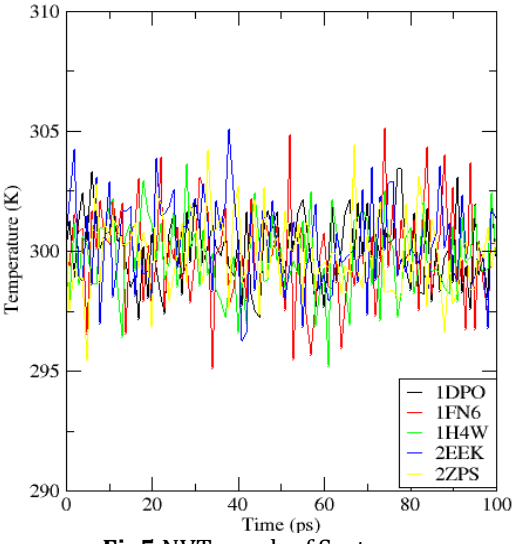

Fig5.NVT graph of Systems

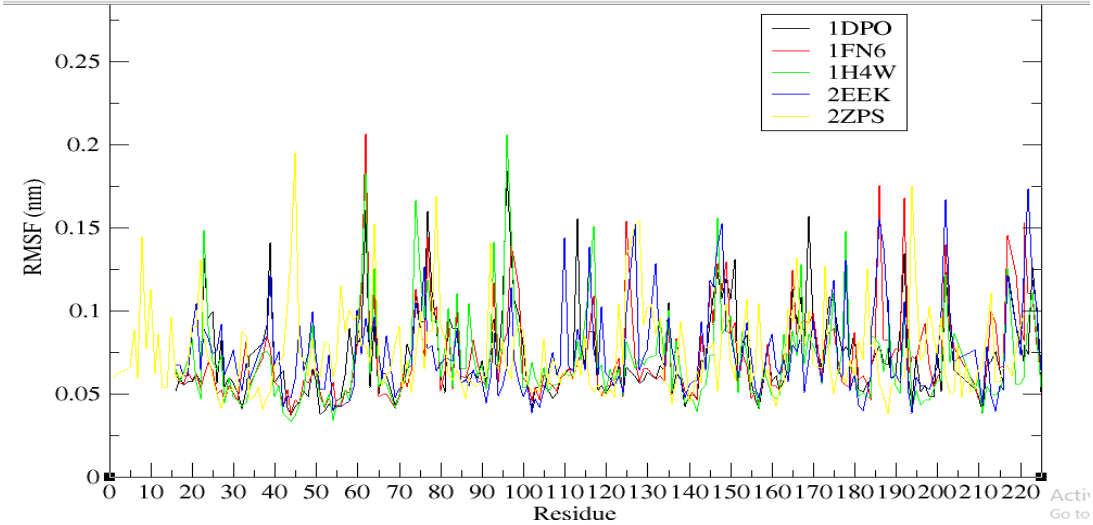

Fig6.RMSF graph of Systems

DISCUSSION

The longest sequence conservation among the sequences under consideration was found at places 34-43 (10 residues), 83-90 (8 residues), 178-185 (8 residues), 15-21 (7 residues), 23-29 (7 residues) (Fig1). When we observe the RMSF at these places, we found that these places have least RMSF values (Fig7). When we observe the highest peaks in RMSF plot, we found that these peaks relates to those regions which possess least or no sequence conservation (Fig7).

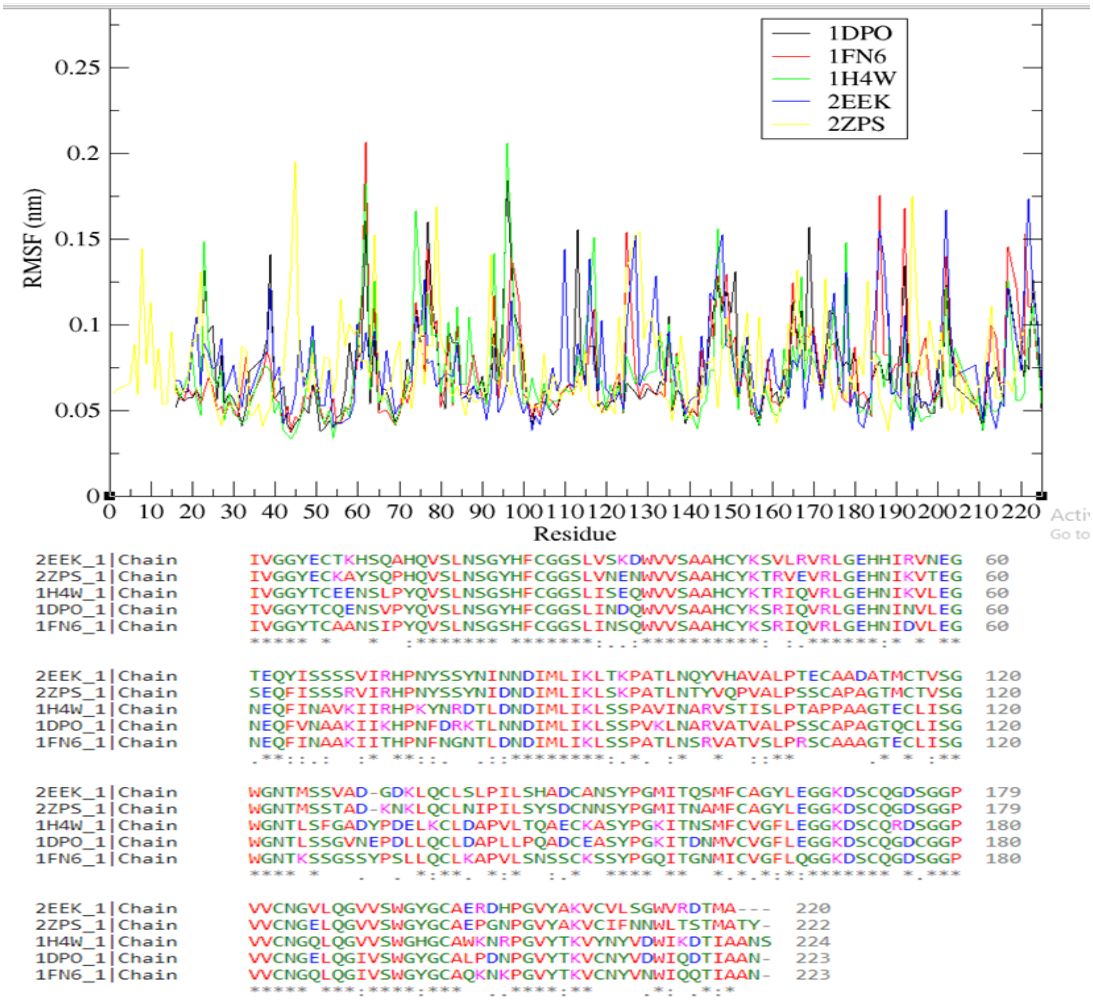

Fig7.RMSF relating with Sequence conservation

CONCLUSION

The aim of this study was to find dynamics properties of amino acids residues of trypsin enzymes. Present work shows that a rational approach based on molecular simulations can successfully be used in studying the dynamic nature of proteins. In this study we found that the places in the trypsin 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.

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.

REFERENCES

1. Webb, E. C. (1992). Enzyme nomenclature 1992. Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes.2 No.Ed.6 pii + 863 pp. ref.5648.
2. Karplus, M. and McCammon, J.A., 1986. The dynamics of proteins. Scientific American, 254(4), pp.42-51.

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