Structural analysis of protein lysyl oxidase: modelling and simulation study

Swechha Mishra*, Puneet Kumar, Sangeeta Singh

Department of Applied Sciences and Bioinformatics, Indian Institute of Information Technology, Allahabad, India.

Lysyl oxidase (Lox) is a regulatory enzyme expressed in various species of lower and higher vertebrate. It remains activated unceasingly as it plays role in interlinking of collagen and elastin via oxidizing lysine residues. Despite its important role, the lysyl oxidase is not well characterized structurally or functionally in humans. 3D model of protein was constructed to analyze its property and function. It was constructed by I-TASSER and a 10 nanosecond (ns) molecular dynamics (MD) simulation was carried out to characterize its structural and dynamical features. Analysis revealed top five modeled structure with varied c-scores of -2.75, -1.98, -3.58, -3.96 respectively. Structure with maximum c-score has ensured the stability of predicted structure after MD simulation.

*Corresponding author: Swechha Mishra, Department of Applied Sciences and Bioinformatics, Indian Institute of Information Technology, Allahabad, India. E-mail: Swechha.iiita@gmail.com

Introduction

Tumor metastasis is a crucial phenomenon from its incipience in primary tumor to blood circulation and advances to other organs [1]. Tumor metastasis may take years to flourish and is one of the major cause of deaths in cancer patients [2]. It is a sequential phenomenon where each step is governed by specific genes such as metastasis initiation, progression and virulency [3]. Tumor metastatic niche is formed by tumor malignant cells and their microenvironment [4], favoring cancer growth. Stromal cells and tumor cells coevolve in the tumor microenvironment and form the permissive niche for cancer cell [5]. Pre-metastasis niche is fathered by soluble factor or microvesicles, released by tumor cells [6]. Tumor

metastasis is frequent in breast, lung, and kidney carcinoma [7, 8]. Hassan Fazilaty reported that high mortality in more than 70% of breast cancer patients is caused by metastasis [9, 10]. Metastasis rate can be abated by apoptosis at various steps as a detachment of tumor cell from the extracellular matrix [11-13]. Lysyl Oxidase gene family encodes five different members, LOX, LOXL1, LOXL2, LOXL3, and LOXL4. These amine oxidases catalyze the covalent crosslink of collagen and elastin side chain component. The C and N terminal of LOX have different activity encoding conserved domain between family members and variable pro-peptide respectively. The other family members of LOX as LOXL2, LOXL3, and LOXL4 contain Cysteine-rich scavenger receptor domain which is supposed to play role in amine oxidase [14]. In view of later

expression, lysyl oxidase (LOX) is thought to play an important role in alternating tumors behavior. It is an extracellular matrix remodeling enzyme which plays a vital role in oxidative deamination. Both tensile strength and structural stability of ECM are control by LOX and further acts to preserve the integrity of tissue [15]. LOX regulates various events in cellular environments such as migration and adhesion. It also regulates compaction of chromatin [16], transcription of gene [17, 18], cellular differentiation, and development of tissue [19].

Materials and methods

Retrieval of protein
Sequence of lysyl oxidase was retrieved from UNIPROT (IDs: Q9Y4KO). This sequence was used as an input for I-Tasser, an open source server for modelling of a protein [20].

Primary structural analysis of Lysyl oxidase
ProtParam from ExPasy (Expert protein analysis system) has been utilized for studying amino acid composition. These Physiochemical characters includes isoelectric point (PI), Molecular weight, molecular formula, instability index, number of positive and negative residues, aliphatic index, extinction coefficient, and grand average hydropathy (GRAVY) [17-22]. Anticipation of Disulphide bond (s_s) was done via CYS_REC tool. The NetPhos-2.0 server was used for studying potential phosphorylation sites of protein [21]. Signal P-4.1 server was used to denote the presence and location of signal peptide cleavage sites in given sequences [22]. At a certain wavelength the amount of light absorbed by protein is defined by extinction coefficient. Proteins in water measured at 280 nm: Molar extinction coefficients (Cys) = 63,315, Molar extinction coefficients (Cys) = 6,260. Thermostability of globular proteins is effected by aliphatic index, and measured by relative volume occupied by aliphatic side chains (Isoleucine, Valine, leucine, and alanine).

Secondary structural analysis of lysyl oxidase protein variants
PSIPRED was used for secondary structure analysis [23]. It is a web based server and works on the principle of two feed - Forward neural network.

Ab Initio modelling, Structure refinement and identification of functional site
Pertaining to lack of a suitable structural template for lysyl oxidase, we resorted to ab initio modeling of lysyl oxidase by using I-TASSER [24]. Initially, PDB library was scanned for possible Profile-Profile threading alignment (PPA). Proteins conformational space was searched by using replica exchange Monte-Carlo simulation, the one with conformational similarities was clustered by SPICKER. The second simulation proceeded via TM-align. One with the lowest energy was chosen after clustering. Finally, Metropolis Monte Carlo simulation was used for docking the domains together which generates the complete model.

TM-score laid between 0 to 1 indicates the significance of the model. When TM-score is ≤ 0.17, it corresponds to the similarity between two randomly chosen PDB libraries. If the TM-score is > 0.5, it corresponds to the similar topology of the structure. Active sites of the modelled protein were retrieved by Cast-P, which predicted the functional site of proteins [25].

Threading and structural association lysyl oxidase
Complete structure of lysyl oxidase was predicted by the iterative implementation of the threading assembly refinement (I-TASSER) server involving two programs: position-specific iterated prediction (PSI-PRED) [26, 27] and position specific iterated-BLAST (PSI-BLAST) [28]. The full-length 3D structure of lysyl oxidase has been generated by I-TASSER servers integrated platform for automated prediction of protein structure and function. The best five structures
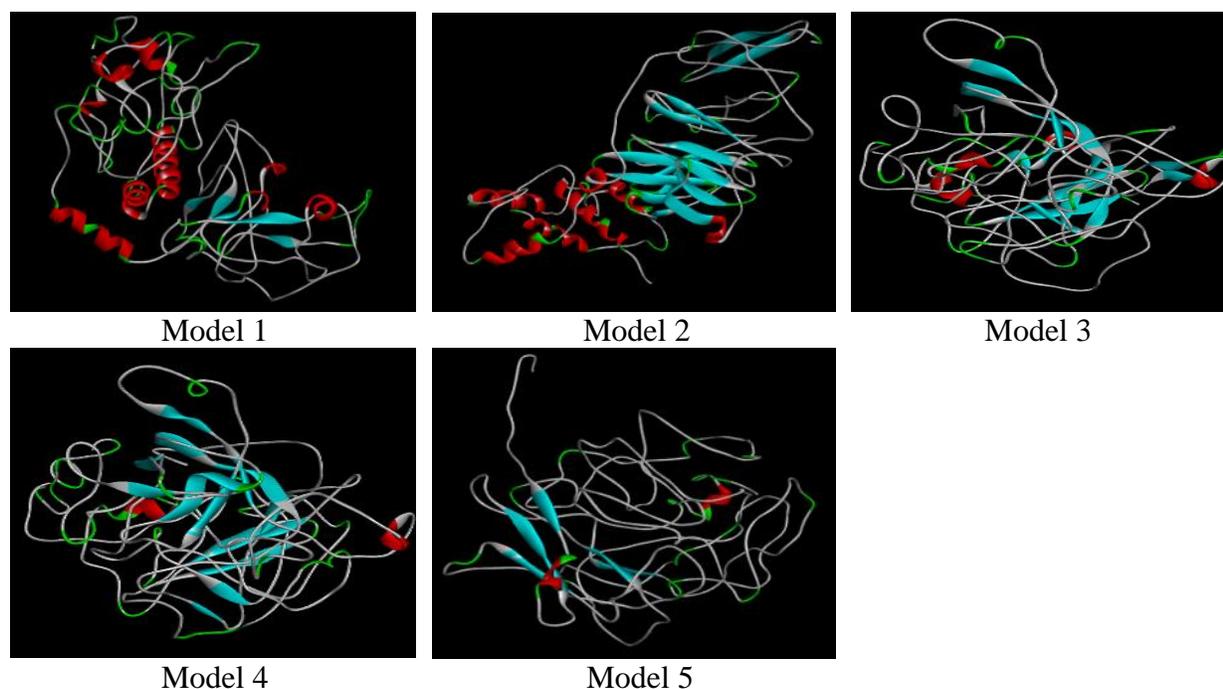
Model 1   Model 2   Model 3

Model 4   Model 5

Figure 1. Predicted model of lysyl oxidase protein. Model 1, 2, 3, 4, and 5 showed different modelled structure of this protein.

obtained from I-TASSER server are shown in figure 1. The high C-score structure was chosen for further validation. Various tools are already imbibed in I-TASSER such as Swiss Model, RaptorX, Modeller, and Phyre, and can be utilized for prediction of the model, peptide structure prediction, and homology modeling [29–32]. LOMETS present in I-TASSER was helpful in examining threading of the modelled protein [33]. It comprises of 10 state-of-art threading programs including SPARKS, HHSEARCH1, COMA, SAM, FUGUE, SP3, MUSTER, PPA1, and PROSPECT2 [34–41]. Threading of predicted secondary structure of query sequence was performed against a PDB library using LOMET server. Top ten templates from each program were chosen for further analysis. ProSa was used for assessment of the quality of generated model [42]. It has given a quality score for the query model in a form of plot which was later analyzed.

Structure Validation of the models obtained by I-TASSER

Qualities of the models were assessed by PSVS and PROCHECK for evaluation and correctness of the overall fold/structure and stereo chemical parameters (angles and bond length). Determination of stereo chemical aspects along with a comprehensive analysis of side and the main chain was completed by validating the modeled structure of the protein. These analyses have confirmed that maximum residues of lysyl oxidase are falling under a permissible region of Ramachandran plot, obtained from PROCHECK.

Molecular Dynamics simulations of lysyl oxidase

Molecular dynamics (MD) simulations, energy minimization, and trajectory analysis were carried out via Desmond. It is a computer-based simulation of the atomic/molecular movements. MD simulation was performed for 4 ns and 10 ns [43, 44]. Neutral territory method was used to explore the parallel computing in simulation process [45]. The interaction between the amino acids in a system was defined by OPLS 2005 force field. To cover the modeled protein structure

SPC (Simple Point Charge) water model was embedded, with predefined dimensions of 10_10_10 [46]. Neutralization of the system was done by adding Cl⁻ ion to maintain the pH of the system. Default steps and protocols of the system were chosen for optimization of the systems equilibrium, allowing them to steady relaxations without deviating from initials co-ordinates. Finally, production run was completed via maintaining the normal temperature and pressure. RMSD and RMSF were calculated for the structural and dynamic behavior of the modelled protein [47].

Results and discussion

Primary structural analysis of lysyl oxidase protein
The primary structural feature of lysyl oxidase is mentioned in table 1. The calculated isoelectric points (pI) for modelled protein is 5.84, which suggested the presence of more negatively charged residues. Modeled protein possesses higher extinction coefficient of 63,315 and 62,690 that indicated the greater presence of Asp, Trp, and Tyr. Two sulfide bonds were depicted in modeled protein. Higher aliphatic index of 51.33 suggested the stability of protein over a wide range of temperature. Instability index of more than 41 suggested thermal instability.

Table 1. Amino acid composition of modelled protein lysyl oxidase.

ProtParam Parameters		
Amino acid composition of lypoxygenase		
Amino Acid	No of Amino Acid	% of Amino Acid
Ala(A)	15	6.0
Arg(R)	18	7.2
Asn(N)	11	4.4
Asp(D)	23	9.2
Cys(C)	10	4.0
Gln(Q)	9	3.6
Glu(E)	8	3.2
Gly(G)	15	6.0

The values of aliphatic index and GRAVY (Grand average hydropathy) 51.33 and -0.848 suggested a hydrophilicity pattern with better interaction with water. Details of all the entity are given in table 2.

Table 2. Physiochemical parameters of modelled protein lysyl oxidase.

No.	ProtParam Parameters	
	Biophysical & Biochemical Properties	*Values*
1	No. of amino acid	417
2	Molecular weight	29031.87
3	Isoelectric point	5.84
4	Negatively charged residues (Asp + Glu)	31
5	Positively charged residues (Arg +Lys)	24
6	Ext. coefficient(C)	63315
7	Abs 0.1 %(C)	2.159
8	Ext. coefficient (CR)	62690
9	Abs 0.1% (CR)	2.181
10	Instability index	44.72
11	Aliphatic index	51.33
12	GRAVY	-0.848

The secondary structure was calculated by PSIPRED-view, which got the highest level of accuracy. An average Q3 score was 72.5%, which was used for the prediction and analysis of the secondary structure of lysyl oxidase proteins, and was shown in table 3.

Table 3. The secondary structure analysis of the modelled protein. The abundance of alpha helix in the protein was 17.51%, followed by other secondary structures of extended strand and random coils which were 20.14% and 54.20% respectively.

Different secondary structure	Percentage (%)
Alpha helix	17.51
3_{10} helix	0
Pi helix	0
Beta bridge	0
Extended strand	20.14
Beta turn	8.15
Bend region	0
Random coil	54.20
Ambiguous states	0
Other states	0

The percentage of helices was 3.93 and that of a strand was 5.07%. The amount of alpha helix and beta sheets was abundant in modelled lysyl oxidase. Procheck validation suite gave an insight into the secondary structure of the protein along with the result of PSI-PRED results. Ramachandran plot of the best model of lysyl oxidase was shown in figure 2. I-TASSER server was used for predicting the 3D structure of lysyl oxidase. Different algorithms of the server generated five models of the target sequence. One with the best C-Score, TM-score, and RMSD values had been chosen in table 4.

Table 4. The three-dimensional analysis of the protein based on quality parameters of C score, T score, and RMSD.

No.	Variants	Score
1	C score	2.75
2	T score	0.40 ± 0.13
3	RMSD	13.6 ± 4.0

The C-score is a confidence score for estimating the quality of predicted models by I-TASSER, which is based on threading template alignments and convergence parameters involved in the simulation. The value of C-Score laid between -3.58 to 1.98 indicated good quality of modeled structure. I-TASSER uses Monte Carlo simulations for reassembling the fragments that are excised at the initial step of the modeling. SPICKER works to cluster these replica structure defined as decoys. Cluster density is constituted by these decoys. A higher value of these signifies the occurrence of structure in simulation trajectory, and thus, indicates the good quality of a model. Structure validation of lysyl oxidase was confirmed via PROCHECK, Verify 3-D, and Molprobity constituted in a single PSVS server. Ramachandran Plot was used for depicting the stereo chemical quality of the lysyl oxidase. A tight clustering of the Phi and Psi bonds was observed. Ramachandran plot basically determines four regions that constrain the amino acids in various allowed and disallowed regions of the plot. Most of the residues (about 90%) were falling into favored regions of the

Ramachandran plot obtained by PROCHECK. Similarly, residues falling in disallowed regions amounted to 8%. This provided an insight into the correctness of the modeled structures in terms of PROCHECK. Comparative study of PROCHECK was performed on selected models after performing the MD simulation study. Tremendous improvement in the model was observed (table. 5).

Table 5. Ramchandran analysis for the validation of modelled protein by using Procheck tools, a comparative analysis of before and after MD simulation study of all the five predicted models.

No.	Model	Most favored (%)	Allowed region (%)	Outliers (%)
1	Pre MD	62.7	21.7	15.7
	Post MD	70.0	19.0	10.0
2	Pre MD	75.2	17.6	7.2
	Post MD	80.0	16.0	4.0
3	Pre MD	60.5	22.5	16.9
	Post MD	68.5	20.3	11.5
4	Pre MD	62.9	24.6	12.5
	Post MD	70.0	20.0	10.0
5	Pre MD	62.4	24.3	13.3
	Post MD	69.5	21.2	12.1

Molecular dynamics simulations of LYSYL OXIDASE protein structures

Desmond (Schrodinger) was used for performing simulation studies initially for 4,000 picoseconds (ps) and further extended to 10,000 ps to have a better insight of stability of modelled protein. The final structures were obtained by taking the average of structures pertaining to stable part of the trajectory. Figures 3 and 4 demonstrate RMSD and RMSF plot of the modelled protein at different time. These structures were subjected for validation by PSVS suite (table 5) for comparison with pre-simulated structure. It is evident from the results that residues falling under most favored regions and allowed regions in Ramachandran plot have significantly increased (PROCHECK and MolProbity) for the modelled protein. The results signify the removal of all the bad contacts from the modelled protein during MD simulation. Similarly, percentage of residues falling in disallowed regions of
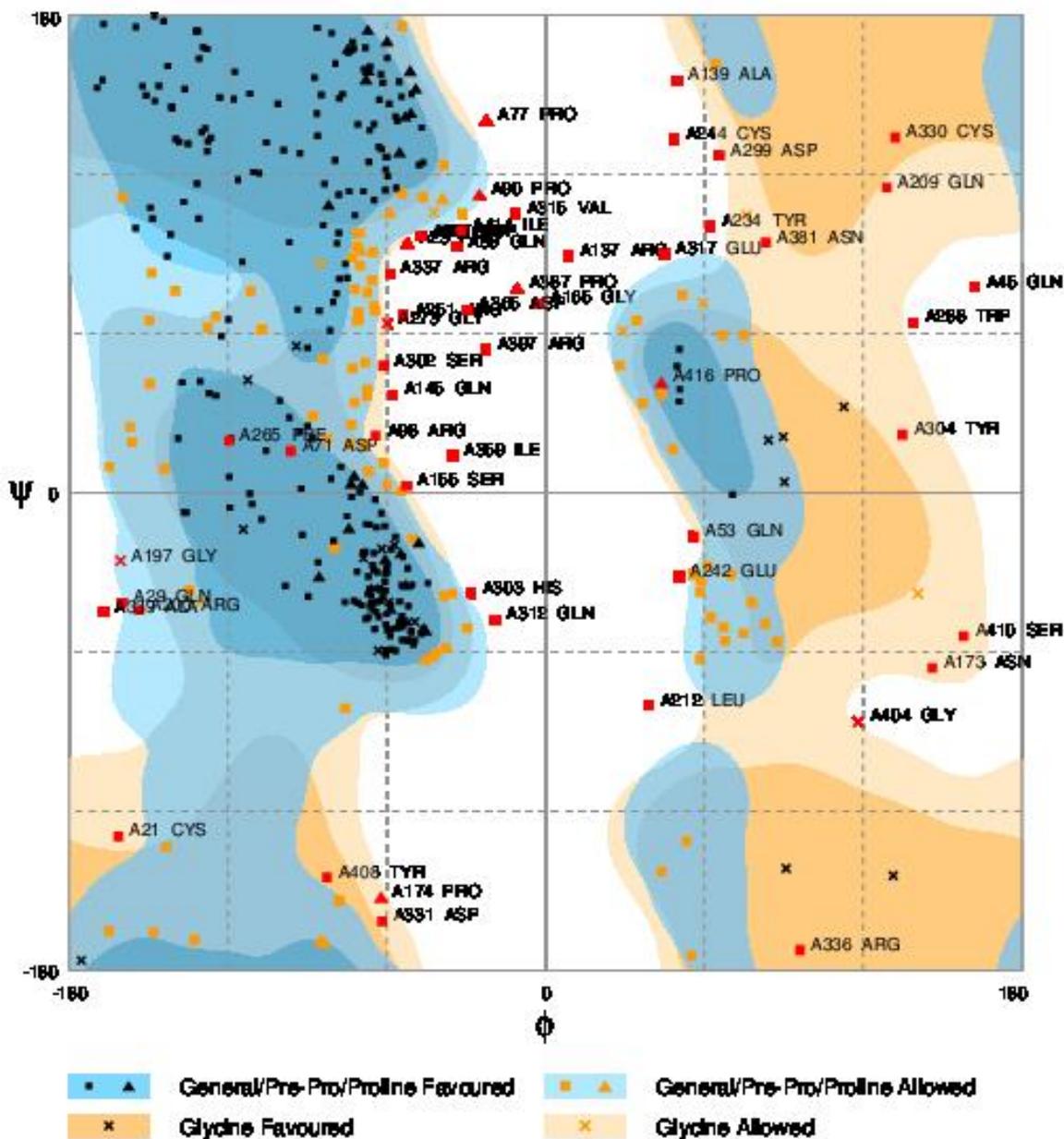
Figure 2. Ramachandran Plot showed the observed data points allocation over the protein, which signified the speculative favored region.

Ramachandran plot decreased by a significant margin leading to an overall improvement in the quality of the models. As mentioned earlier, it consists of single α-helix hairpin chain with 417 amino acids. These loops can be result of various stereo chemical constraints under which a complete protein is modelled. In absence of other chains, algorithm takes the single fragment into perspective and puts stereo

chemical constraints occurring within that part. The top template selected for modelling was crystal structure of human 4bedB with a Z score of 0.887. The Z score was given by:

$$Z_{score} = (R'_{score} - (R'_{score})) / ((R'^2_{score}) - (R'_{score}) 2)^{1/2}$$

where R'_{score} was a normalized score and was derived by dividing raw alignment score (R_{score})
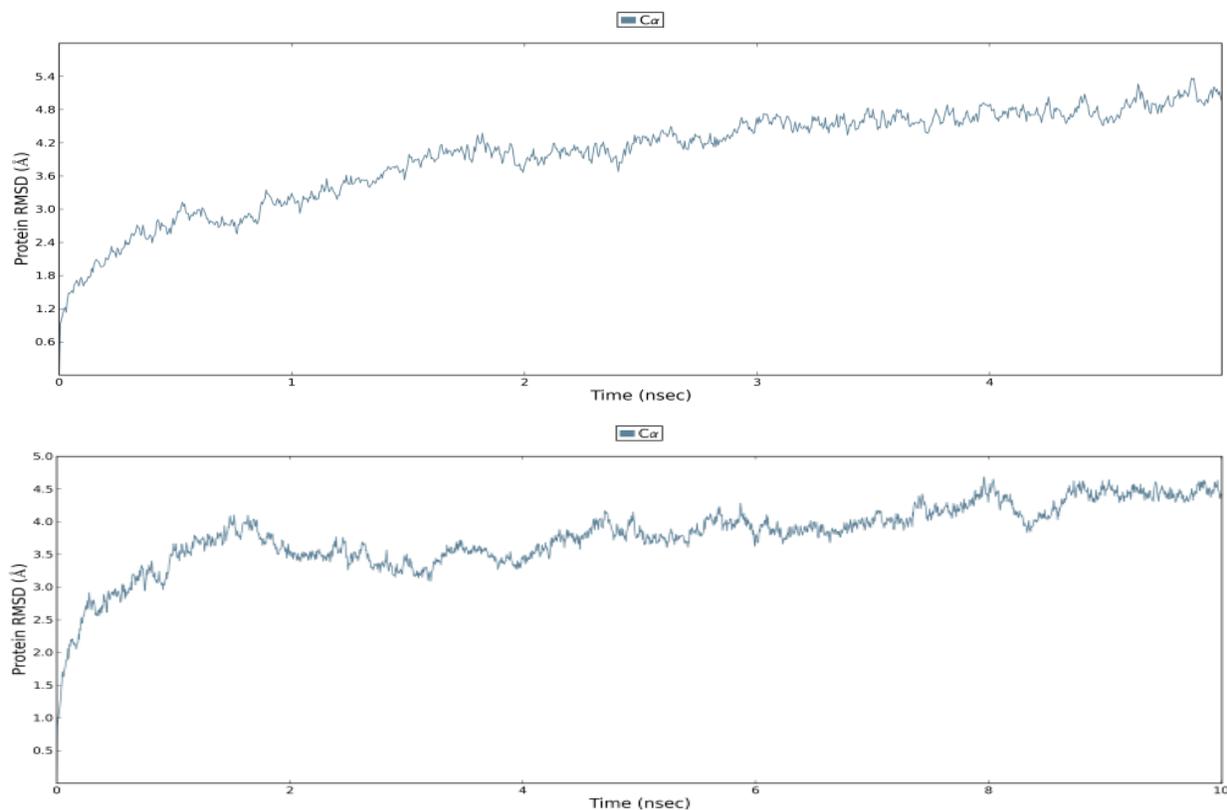
Figure 3. RMSD of modelled protein to quantify the resemblance among overlaid atomic coordinates at 4 ns and 10 ns. RMSD is the middling over atoms with respect to time when calculated over a trail.

with that of partial alignment length (L$_{partial}$ i.e. excluding query ending gap). During the whole MD, simulation protein was found to be very stable, thus validated the findings that they existed as a single chain structure and were stable in their conformation. Obtained model provided crucial insight into topological specifications that were responsible for the functionality of the protein. Making as much use of available techniques, we attempted to identify the protein conformations the closest to its biologically relevant conformations.

Conclusion

Lysyl oxidase play vital role in various metabolic mechanisms. It also plays significant role in controlling metastasis. In various experimental studies upregulation of lysyl oxidase has been found. Researchers are also studying it as a new target against metastasis. Being a small protein and owing to sterical hindrance, it is difficult to isolate it, thus posing string hindrances in experimental structural analysis. In light of this computational approach has become viable to obtain 3D structure, it leads to detect such protein's functional and structural relationship. Analysis has revealed that protein is of 417 amino acids and consist of one alpha chain and having coiled configuration with loops, which make them flexible. Estimation of various physiological and biochemical properties has facilitated in understanding mechanism of its action. This study opens a way for further attempts to elucidate the experimental structures and validating the accuracy of the predictions. Structural knowledge of the protein is important for deducing its function while interacting with various substrates such as

Figure 4. RMSF plot of modelled protein for 4 ns and 10 ns. Greater stability can be seen at 10 ns.

primary amine. Structural information is required for understanding and assigning the functional role to the protein.

Reference

1. Elizabeth Ward, Carol DeSantis, Anthony Robbins, Betsy Kohler, Ahmedin Jemal. 2014. Childhood and Adolescent. Cancer Statistics. 5:1-21.

2. Nguyen DX, Bos PD, Massague J. 2009. Metastasis: From dissemination to organ-specific colonization. Nat. Rev. Cancer. 9:274–284.

3. Chiang AC & Massague J. 2008. Molecular basis of metastasis. N. Engl. J. Med. 359:2814–2823.

4. Scadden DT. 2006. The stem-cell niche as an entity of action. Nature. 441:1075–1079.

5. Quail DF, Joyce JA. 2013. Microenvironmental regulation of tumor progression and metastasis. Nat. Med. 19:45-59.

6. Peinado H et al. 2012. Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET. Nat. Med. 18:883–891.

7. Shimaoka K, Sokal JE, Pickren JW. 1962. Metastatic neoplasms in the thyroid gland. Pathological and clinical findings. Cancer. 15:557–565.

8. Wychulis AR, Beahrs OH, Woolner LB. 1964. Metastasis of carcinoma to the thyroid gland. Annals of Surgery. 160:169–177.

9. Mundy GR. 2002. Metastasis to bone: causes, consequences, and therapeutic opportunities. Nat. Rev. Cancer. 2:584–593.

10. Roodman GD. 2004. Mechanisms of bone metastasis. N Engl. J. Med. 350:1655–1664.

11. Wyckoff JB, Jones JG, Condeelis JS, Segall JE. 2000. A critical step in metastasis: in vivo analysis of intravasation at the primary tumor. Cancer Res. 60:2504–2511.

12. Wong CW, Lee A, Shientag L, Yu J, Dong Y, Kao G et al. 2001. Apoptosis: an early event in metastatic inefficiency. Cancer Res. 61:333–338.

13. Martin SS, Vuori K. 2004. Regulation of Bcl-2 proteins during anoikis and amorphosis. Biochim Biophys Acta. 1692:145–157.

14. Kim MS, Kim SS, Jung ST, Park JY, Yoo HW, Ko J, Csiszar K, Choi SY, and Kim Y. 2003. Expression and purification of enzymatically active forms of the human lysyl oxidase-like protein 4. J. Biol. Chem. 278:52071–52074.

15. Kagan HM, Li W. 2003. Lysyl oxidase: properties, specificity, and biological roles inside and outside of the cell. J. Cell. Biochem. 88:660–672.

16. Mello ML, Contente S, Vidal BC, Planding W, Schenck U. 1995. Modulation of ras transformation affecting chromatin supraorganization as assessed by image analysis. Exp Cell Res. 220:374–382.

17. Di Donato A, Lacal JC, Di Duca M, Giampuzzi M, Ghiggeri G et al. 1997. Micro-injection of recombinant lysyl oxidase blocks oncogenic p21-Ha-Ras and progesterone effects on Xenopus laevis oocyte maturation. FEBS Lett. 419:63–68.

18. Giampuzzi M, Botti G, Di Duca M, Arata L, Ghiggeri G et al. 2000. Lysyl oxidase activates the transcription activity of human collagen III promoter. Possible involvement of Ku antigen. J Biol Chem. 275: 36341–36349.

19. Maki JM, Rasanen J, Tikkanen H, Sormunen R, Makikallio K et al. 2002. Inactivation of the lysyl oxidase gene Lox leads to aortic aneurysms, cardiovascular dysfunction, and perinatal death in mice. Circulation. 106: 2503–2509.

20. Zhang Y. 2008. I-TASSER server for protein 3D structure Prediction. BMC Bioinformatics. 9:40-45.

21. Petersen TN, S Brunak, G Heijne, H Nielsen. 2011. SignalP 4.0: Discriminating Signal Peptides from Transmembrane Regions. Nature Methods. 8:785-786.

22. Gill SC, PH Hippel. 1989. Calculation of Protein Extinction Coefficients from Amino Acid Sequence Data. Anal Biochem. 182:319- 326.

23. McGuffin LJ, Bryson K, Jones DT.2000.The PSIPRED protein structure prediction server. Bioinform Appl Note.16 (4):404–405.

24. Roy A, Kucukural A, and Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc. 5 (4):725–738.

25. Laurie ATR, Jackson RM. 2005. Q-Site Finder: an energy-based method for the prediction of protein–ligand binding sites. Bioinformatics. 21(9):1908–1916.

26. McGuffin LJ, Bryson K, Jones DT. 2000. Bioinformatics. 16:404–405.

27. Jones DT. 1999. Protein secondary structure prediction based on position-specific scoring matrices. Journal of molecular biology. 292(2):195-202.

28. Margelevičius M and Venclovas Č. 2005. PSI-BLAST-ISS: an intermediate sequence search tool for estimation of the position-specific alignment reliability. BMC bioinformatics. 6(1):1-20.

29. Eswar N, Marti-Renom MA, Webb B, Madhusudhan MS, Eramian D, Shen M, Pieper U, Sali A. 2006. Current Protocols in Bioinformatics.Wiley. New York. Supplement 15:(5)6.1–6.30.

30. Kelley LA and Sternberg MJ. 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nature protocols. 4 (3):363-371.

31. Xu J, Li M, Kim D and Xu Y. 2003. RAPTOR: optimal protein threading by linear programming. Journal of bioinformatics and computational biology. 1 (1):95-117.

32. Schwede T, Kopp J, Guex N and Peitsch MC. 2003. SWISS-MODEL: an automated protein homology-modeling server. Nucleic acids research. 31(13):3381-3385.

33. Wu S and Zhang Y. 2007. LOMETS: alocal meta-threading-server for protein structure prediction. Nucleic acids research. 35(10):3375-3382.

34. Margelevičius M, Laganeckas M and Venclovas Č. 2010. COMA server for protein distant homology search. Bioinformatics. 26(15):1905-1906.

35. Shi J, Blundell TL and Mizuguchi K. 2001. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. Journal of molecular biology. 310(1):243-257.

36. Söding J. 2005. Protein homology detection by HMM–HMM comparison. Bioinformatics. 21(7):951-960.

37. Wu S and Zhang Y. 2008. MUSTER: improving protein sequence profile–profile alignments by using multiple sources of structure information. Proteins: Structure, Function, and Bioinformatics. 72 (2):547-556.

38. Xu Y and Xu D. 2000. Protein threading using PROSPECT: design and evaluation. Proteins: Structure, Function, and Bioinformatics. 40(3):343-354.

39. Karplus K, Karchin R, Draper J, Casper J, Mandel-Gutfreund Y, Diekhans M and Hughey R. 2003. Combining local-structure, fold-recognition, and new fold methods for protein structure prediction. Proteins: Structure, Function, and Bioinformatics. 53(6):491-496.

40. Zhou H and Zhou Y. 2004. Single-body residue-level knowledge-based energy score combined with sequence-profile and secondary structure information for fold recognition. Proteins: Structure, Function, and Bioinformatics. 55(4):1005-1013.

41. Zhou H and Zhou Y. 2005. Fold recognition by combining sequence profiles derived from evolution and from depth-dependent structural alignment of fragments. Proteins: Structure, Function, and Bioinformatics. 58 (2):321-328.

42. Wiederstein M and Sippl MJ. 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic acids research. 35(2):407-410.

43. Kevin JB, Edmond C, Huafeng X, Ron OD, Michael PE, Brent AG, John LK, Istvan K, Mark AM, Federico DS, John KS, Yibing S, David ES. 2006. Scalable algorithms for molecular dynamics simulations on commodity clusters. In: Proceedings of the ACM/IEEE conference on supercomputing, ACM, Tampa.

44. Purohit R. 2014. Role of ELA region in auto-activation of mutant KIT receptor: a molecular dynamics simulation insight. J Biomol Struct Dyn. 32(7):1033–1046.

45. Bowers KJ, Dror RO, Shaw DE. 2006. The midpoint method for parallelization of particle simulations. J Chem Phys. 124:184109–184111.

46. Jorgensen WL, Maxwell DS, Tirado-Rives J. 1996. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc. 118:11225–11236.

47. Kumar A, Purohit R. 2014. Use of long term molecular dynamics simulation in predicting cancer associated SNPs. PLoS Comput Biol. 10(4):1003310-1003318.