© 2016 IJSRST | Volume 2 | Issue 4 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

Bioinformatics Analysis of Evolution of Secondary Structures of Protein Trypsin Beta

Mahin Ghorbani*

Department of Biotechnology, Fergusson College, F. C. Road, Pune, Maharashtra, India

ABSTRACT

Impact of divergence of amino acid sequences of protein during evolution on the secondary structures of polypeptide and phylogenetic significance of different secondary structures such as alpha helix, beta sheet and random coil in evolution are not known. Due to conservation of conformation ally identifiable regions through evolution, in closely related species, the amino acid sequence will show close similarity, thereby giving rise to similar structural motifs after folding the proteins. To understand the most conserved secondary structure element of a protein, we have conducted a bioinformatics work for molecular evolution of protein trypsin beta as a sample in order to analysis the phylogeny of secondary strucutres (alpha helix, beta sheet and random coil) of proteins Trypsin Beta among 25 species. In this method we retrieved amino acid sequences of these proteins from 25 species from protein data bank then folded each individually into 3-D structure using the software J-Pred. From the folded sequence it was possible to identify sequences in regions forming alpha helix, Beta sheet, random coil, which we retrieved and individually ligated end-end to obtain peptides made up of sequence in the random coil, alpha helix and beta sheet conformations (final functional shape). Then examined the phylogenetic trees built after aligning the sequence using four different multiple alignment protocols. The result has assumed that random coil of Trypsin beta was phylogentically most conserved. This project plays significant role in understanding the role of molecular evolution of proteins and their phylogenic significance.

Keywords: Bioinformatics, Molecular Evolution, Secondary Structures, Trypsin Beta, Multiple Sequence Alignment, Phylogeny.

I. INTRODUCTION

Proteins are sequences of amino acids which form primary structure of the poly peptides. [1]. To become active , a protein should be folded , this folding initially started in the form of secondary structures which finally led into three dimensional form of the protein. The secondary structure elements of the proteins are called, Alpha helix, Beta sheet and random coil. [1]. Description of Protein secondary structure focused on the pattern of the hydrogen-bonding of the peptide backbone of the protein. [1], [2], The alpha helix (α-helix) is the most common, regular segment of the secondary structure of proteins. [1, 2]This right-handed coil shows donation of a hydrogen bone from backbone of N-H group to the backbone C=O group of the amino acid four residue earlier. It is also called as 3. 613-helix , denoted the number of residues per turn, and involvement of 13 atoms in the ring formed by the hydrogen bond. [1, 2]The β sheet (also β-pleated sheet) is the second element of secondary structures in proteins. [1, 2]. As compared to alpha helix, it is less common. Beta sheets formed of beta strands joined laterally by minimum two or three backbone hydrogen bonds, which led in formation of twisted pleated sheets. A beta strand is typically 3 to 10 amino acids long which formed a stretch of polypeptide. The beta sheet exhibits in two forms parallel and anti-parallel. [1, 2] The third element contributed in secondary structure of protein is called random coil, which is a polymer and it's monomer unit's orientation take place randomly. It is a major portion of the protein which lies at the surface of the protein. In respect to the significance in function, alpha helix plays the most important role and beta sheet plays scaffold role. [1, 2]Secondary structures serve an important role in stabilizing of the overall folding of protein by

providing the high amount of the enthalpy of stabilization of folding which led in existence of polar backbone groups in the hydrophobic of a folded polypeptide. [3] Divergence of protein sequences occurred during evolution, [4] but it is not known how this divergence impact secondary structures segments, nor significance of phylogenetic status of the secondary structure members in evolution. For instance, some of conformationally detectable regions shows conservation throughout evolution, so that , in closely related species , the amino acid sequences will show high degree of similarity thereby giving rise to similar structural motifs after folding the proteins. [4]

To find out the degree of conservation of secondary structure elements and phylogenetics status of individual element in order to analysis their significance in evolution , we have selected the trypsin beta protein as material and performed it's bioinformatics analysis in order to introduce this novel method for further bioinformatics analysis of different secondary structures of other important proteins and their role in evolution and other evolution related issues relevant to the secondary structures of the proteins.

II. METHODS AND MATERIAL

we have selected 25 species having Trypsin beta such as Rarobacter faecitabidus, Boltenia villosa, Bos Taurus, Canis lupus familiaris , Rattus norvegicus, Mus musculus , Homospaiens, Macaca mulatta, Trimeresurus jerdonii, Bitis gabonica, Anplopoma fimbria, Gadus morhua, Salmo salar, Tryophagus putrscentiae, Lonomia obliqua , Mamestra configurata, Hypoderma Diana, Drosophila virillis, Drosophila melanogaster, Phlebotomus papatasi, Lutzomyia longipalpis, Anopheles stephensi, Cultex quinquefasciatus, Caligus rogercresseyi, Radix peregra , Loligo bleekeri , from Protein data bank and downloaded amino acids sequences of them in FASTA format. Then using online primary structure to secondary structure conversion software j–pred , [5] by which the amino acid sequences of proteins of each species converted into potential secondary structures polypeptide. The converted folded polypeptide (contain all elements such as alpha helix , beta sheet and random coil) of all 25 species cut off into the secondary structure elements serially and then individually ligated end-end to obtain peptides made up of sequence in the random coil, alpha helix and beta sheet conformations. Then these insilco ligated polypeptide along with original one subjected to four multiple sequence alignments namely:

1. Clustal w: (http://www. ch. embnet. org/software/ClustalW. html) [6] 2. ProbCon, (http://toolkit. tuebingen. mpg. de/probcons) [7], 3. Tcoffee, (http://toolkit. tuebingen. mpg. de/t_coffee[8] 4. Mafft(http://www. ebi. ac. uk/Tools/msa/mafft/)[9]

The results of all multiple sequence alignment saved in separate notepads. , then the result of each multiple sequence alignment softwares were subjected in Protdist software, (http://caps. ncbs. res. in/iws/protdist. html)[10] then the out file result of protdist subjected in neighbor software, [11] then the neighbor outtree ‘s result saved in a separate notepad, then the outtree's result opened with MEGA, [12]then species clustered using tree view X software(http://darwin. zoology. gla. ac. uk/~rpage/treeviewx/)[13]. and saving in separate notepad and rename it as cluster file. txt, then outtree result of each software saved in separate notepad and rename as outtree. txt, then , then we opened python cluster software and enter in first line cluster. txt and in second line outtree. txt and click on enter the result which will appear automatically in a notepad and open it with Excel software, then calculation of the percentage of fidelity of each secondary structure obtained and saved in Excel file.

III. RESULT AND DISCUSSION

For each MSA methods, different percentages of secondary structures conservation obtained, based on their specific performances. **Table 1**. Molecular evolution analysis of secondary structure (random coil, alpha helix Beta sheet) of Trypsin Beta protein by 4 different MSA is discussed as below: According to the result obtained from clustal w: The most conserved secondary structure is alpha helix and the second one is random and the least conserved(The most varied) is Beta sheet , which indicates that the most mutation occurred in Beta sheet during course of molecular evolution. The conservation percentages obtained from clustalw for Trypsin protein‘s secondary structure is as below: Alpha helix >Random coil> Beta sheet (61. 78%>59. 17%>55. 47%respectively). On the basis of the result obtained from Mafft method: The most conserved part of secondary structure forTrypsin Beta is Random

coil and the second one is Beta sheet and the least one is alpha helix as below: Random coil >Beta sheet> Alpha helix (57. 17%>55. 73%>53. 78%respectively). The highest percentage is random coil which shows the most conservation. On the basis of the result obtained from ProbCons method: The most conserved part of secondary structure is random coil and the second one is beta sheet and the least one is alpha helix. The conservation percentages obtained from probcons for Trypsin protein ‘s secondary structure is as below:Random coil > Beta sheet>Alpha helix (61. 08 %>55. 04% >54. 6 %respectively). The highest percentage is random coil which shows the most conservation. On the basis of the result obtained from Tcoffee Method: The most conserved part of secondary structure is Random coil and the second is Alpha helix and the least is Beat sheet. The conservation percentages obtained from Tcoffee A for protein TIFIIA‘s secondary structure is as below: Random coil >Alpha helix > Beta sheet 60. 95%>56. 47%>55. 82% respectively). The highest percentage is random coil which show the most conservation. Overall results of these finding is considered carefully and due to negotiable difference between percentages in different MSA, the result has shown that random coil phylogentically is the most conserved part of the protein

Table 1: Percentage of conservation of secondary structures of Trypsin beta by four different multiple sequence alignment protocol**:**

Multiple sequence alignment	Random coil %	Alpha Helix %	Beta sheet %
Clustal w	59. 17	61. 78	55. 47
Mafft	57. 17	53. 78	55. 73
Probcons	61. 08	54. 6	55. 04
Tcoffee	60. 95	56. 47	55. 82

Phylogentic trees of secondary structures of trypsin beta using four different multiple sequence alignments constructed and compared with benchmark tree of classification of 25 species and visual analysis of the trees showed that in the tree constructed using clustal w , beta sheet element of mammalian underwent mutation and human's beta sheet of trypsin beta is diverged from other four mammalian speices. For the nine studied fly species , the beta sheet is scattered and three of the fly species come together and remaining six are evolved together, also in this tree, tunicate and crustacean species evolved together , bacteria specie come near by two mollusca species but two snake species and three bony fishes species evolved tightly together. For random coil flies species are scattered in four regions and mammalian species split into two places and two mollusca underwent divergence but still snake and bony fishes are evolved together tightly which showed high degree of conservation of the element among these species. and for alpha helix the conservation degree showed maximum amount as there is no divergence of this segment among mammalian species , snake and bony fishes species but still divergence of one flies species is shown , so visually alpha helix element showed the maximum degree of conservation and beta sheet most mutated one and random coil takes place an intermediate place.

III. Discussion

Conversion of linear polymer of polypeptide into a stable three dimensional functional protein is a thermodynamically dependent process [13]. secondary structures formation occurred during initial stages of protein folding so their role in protein folding is crucial as they provide much of enthalpy for stabilization of protein during folding. [14] As the secondary structures play significant role in protein folding, their conservation, evolution and phylogenetic status are significantly considered for understanding the entire protein evolution and mutation and other phylogenetic ally related issues. Here we have conducted an original work to find out the molecular evolution of secondary structure of trypsin beta. In this work, we have examined performance of four different multiple sequence alignments and for each MSA different percentages are obtained. Due to negotiable difference between percentages obtained from different multiple sequence alignment, the result has shown that random coil phylogentically is the most conserved. On the basis of phylogentic analysis of trees constructed based on secondary structures and compared with benchmark trees in which the species clustered in accordance with their taxonomy, it visualizes scattered distribution of

species and their genus , and evidences based on the trees shows that evolution of secondary structures occurred many times , although the exact reason of evolution is not known but occurrence of evolution within the secondary structures are well observed by this method.

IV. CONCLUSION

This work is helpful in understanding how secondary structures of a protein diverge or evolve among closely related species and the most conserved part of secondary structures of a protein is calculated by multiple sequence alignment protocol. Although still the reasons for evolution of secondary structures are open for discussion but this work is a starting point to study the reason of the molecular evolution and their significant role in the evolution and function of the appropriate protein. This projct can be helpful for understanding protein structure and its evolution in drug discovery projects as many targets are protein. For more information about protein structure of drug targets please referee to the following publications.[15-20].

V. REFERENCES

[1] Drenth, J.: Principles of Protein X-Ray Crystallography 2nd ed. (Springer-Verlag,New York, 1999).

[2] Evans, J.N.S.: Biomolecular NMR Spectroscopy (Oxford University Press,Oxford, 1995).

[3] Schmidt, A. and Lamzin, V.S.: Veni, vidi, vici—atomic resolution unravelling the mysteries of protein function. Curr.Opin. Struct. Biol. 2002,12:698–703.

[4] Brooks DJ, Fresco JR, Lesk AM, Singh M.Evolution of amino acid frequencies in proteins over deep time: inferred order of introduction of amino acids into the genetic code. Mol Biol Evol. 2002 Oct;19(10):1645-55.

[5] Cole C, Barber JD & Barton GJ. Nucleic Acids Res. 2008. 35 (suppl. 2) W197-W201 link]

[6] Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J. and Higgins D.G. (2007)ClustalW and ClustalX version 2.Bioinformatics 23(21): 2947-2948

[7] Do CB, Mahabhashyam MS, Brudno M, Batzoglou S. ProbCons: Probabilistic consistency-based multiple sequence alignment. Genome Res. 2005 Feb;15(2):330-40.

[8] Magis C, Taly JF, Bussotti G, Chang JM, Di Tommaso P, Erb I, Espinosa-Carrasco J, Notredame C. T-Coffee: Tree-based consistency objective function for alignment evaluation. Methods Mol Biol. 2014;1079:117-29. doi: 10.1007/978-1-62703-646-7_7.

[9] Katoh, Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. (Outlines version 7) 2013 (Molecular Biology and Evolution 30:772-780

[10] Felsenstein, J. 1989. PHYLIP -- Phylogeny Inference Package (Version 3.2). Cladistics 5: 164-166

[11] Wheeler, T.J. 2009. Large-scale neighbor-joining with NINJA. In S.L. Salzberg and T. Warnow (Eds.), Proceedings of the 9th Workshop on Algorithms in Bioinformatics. WABI 2009, pp. 375-389. Springer, Berlin.

[12] Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei (2001) MEGA2 : Molecular Evolutionary Genetics Analysis. Ver. 2.0, Bioinformatics 17:1244-1245

[13] Poland D. Contribution of secondary structure to the heat capacity and enthalpy distribution of the unfolded state in proteins, Biopolymers. 2002 Jan;63(1):59-65.

[14] Sathyanarayana N. Gummadi.What Is the Role of Thermodynamics on Protein Stability, Biotechnology and Bioprocess Engineering 2003, 8: 9-18

[15] Ghorbani M and Karimi H. Cyclin-Dependent Kinases as valid targets for cancer treatment. Journal of Pharmacy Research 2015,9(6),377-382

[16] Ghorbani M, Karimi H, 'Ion Channels Association with Diseases and their Role as Therapeutic Targets in Drug Discovery', International Journal of Scientific Research in Science and Technology(IJSRST), 1(3):65-69,July-August 2015.

[17] Ghorbani M, Karimi H, 'Role of Aquaporins in Diseases and Drug Discovery', International Journal of Scientific Research in Science and Technology(IJSRST),1(3):60-64, July-August 2015

[18] Mahin Ghorbani, Hamed Karimi, 'Role of G-Protein Coupled Receptors in Cancer Research and Drug Discovery', International Journal of Scientific Research in Science and Technology (IJSRST),1(3), pp.122-126, July-August 2015.

[19] Mahin Ghorbani, Hamed karimi, 'Role of Biomarkers in Cancer Research and Drug Development', International Journal of Scientific Research in Science and Technology(IJSRST),1(3), pp.127-132, July- August 2015

[20] Mahin Ghorbani, Hamed Karimi, 'Bioinformatics Methods for Biochemical Pathways and System Biology Analysis', International Journal of Scientific Research in Science and Technology (IJSRST),1(4),75-79, September-October 2015.