Original Research Article **Bioinformatics**

International Journal of Pharma and Bio Sciences

ISSN 0975-6299

INSILICO ANALYSIS OF PB1 PROTEIN OF INFLUENZA A VIRUS

K.AKILA

PG and Research Department of Biotechnology and Bioinformatics,
Bishop Heber College, Tiruchirappalli-620 017.

ABSTRACT

The *influenza A* virus affected a worldwide pandemic outbreak in 2009. Due to the reoccurrence of H1N1 pandemic there is a necessity to study the H1N1 proteins. Influenza A virus is a lipid-enveloped orthomyxovirus and a origin of human disease by strains that arise through periodic variation and through pandemic infection causing from viral adaptation that introduces new influenza viruses into the human population. Similar other influenza viruses cause a respiratory infection that spreads from individual to individual by coughing and sneezing. The current study focuses PB1 protein of the *Influenza A* virus. This protein is the subunit of polymerase protein. It is involved in both transcription and replication of the genome RNA. It has 757 amino acids. This work involves the targeting of valine at 715th position in the protein sequence which is the most mutable site in the sequence as inferred from the literature. The site was replaced by hydrophilic amino acid in the 715th position of the sequence instead of hydrophilic amino acid. Due to the most harmful effect of the mutated protein, the current work aimed at designing the 3D structure and the predicted structure was validated. This modelled structure will aid to design a drug in future.

KEYWORDS: ***Influenza A, PB1, Mutational Analysis, Designing.***

K.AKILA

PG and Research Department of Biotechnology and Bioinformatics,
Bishop Heber College, Tiruchirappalli-620 017.

Received on: 30-01-2017
Revised and Accepted on: 16-05-2017
DOI: http://dx.doi.org/10.22376/ijpbs.2017.8.3.b140-144

INTRODUCTION

Influenza *A* virus causes influenza in birds and some mammals. *Influenza* virus A is a species of the *Orthomyxoviridae* family of viruses. Influenza virus responsible for causing a global endemic was originated from Mexico followed by USA. In 2009, a novel strain of H1N1 influenza virus emerged in California and rapidly spread throughout the world[13]. A recent study estimated that >284,000 deaths occurred globally during the first 12 months of 2009 pandemic H1N1virus circulation[6]. Since last two years, there is gradual increase in the reported case of H1N1 infection and according to World Health Organization (WHO) report, influenza activity increased in several areas of the Southern Hemisphere which is dominated by the H1N1 pandemic strain of 2009[2]. According to Union Health Ministry, India the number of affected people across the country has mounted to 25,190 with 1370 death till March, 2015. Due to the reoccurrence of H1N1 pandemic there is a need to study the evolution of different H1N1 proteins so as to make out spread of the disease. Influenza A virus is a lipid-enveloped orthomyxovirus and a cause of human disease by strains that arise through seasonal variation and through pandemic infection resulting from viral adaptation that introduces new influenza viruses into the human population[4]. Influenza A virus strains are assigned an H and an N number based on these two proteins, the strain contains. There are 16 H and 9 N subtypes known in birds, but only H 1, 2 and 3, and N 1 and 2 are commonly found in humans. Influenza A (H1N1) virus is a subtype of influenza virus A and the most common cause of influenza (flu) in humans. The genetic material of Influenza virus contains eight segments of single stranded RNA. Out of its 8 segments of RNAs, 2 polymerase genes, PB2 and PA, were from the avian virus of North American lineage and were introduced into swine populations around 1998. The other polymerase gene, PB1, also evolved recently from a human seasonal influenza (H3N2) virus around the same year. Hemagglutinin (HA), nucleoprotein (NP) and nonstructural (NS) protein coding genes descended directly from the classic swine influenza A virus of North American lineage, which can be traced back to the 1918 virus. The matrix (M1) protein of influenza A virus is a multifunctional protein that plays essential structural and functional roles in the virus life cycle[10]. HA (hemagglutinin) has been demonstrated to be particularly important for virus infection against the host, by mediating the attachment of the virus to the host cell surface and the entry of viral RNA into the host cell. Therefore, the properties of the HA protein in H1N1 virus are very worthwhile to be studied, which will provide a clue to better understand the infection mechanism of influenza viruses and monitor the interspecies transmission of influenza virus[8]. HA (hemagglutinin) and NA (neuraminidase) play roles in viral attachment and release from host cells, respectively[5]. The nonstructural gene (NS) of the influenza A virus has a crucial role in viral virulence and replication[17]. *Influenza A* viruses are pandemic due to sudden mutation/variation in surface proteins. There are records of evidence that the *Influenza A* virus may mutate into a form that can be transmitted to human easily. The mutations lead to different forms of surface proteins that form different structure. This study focuses PB1 protein of the *Influenza A* virus. This protein is the subunit of polymerase protein. It is involved in both transcription and replication of the genome RNA. It has 757 amino acids. In earlier mutation is studied in PB1 protein using self-made program CARd. This program helps in identification of ligand binding sites and to understand the problem of protein and protein-DNA interactions[1]. The mutation is carried out using hydrophilic amino acid in the 715 position of the sequence instead of hydrophilic amino acid. The mutated protein is subjected to design the 3D structure and the predicted structure was validated.

MATERIALS AND METHODS

Sequence retrieval

The sequence of the protein PB1 has been retrieved from the NCBI database. Its accession number is NP_040985.1.Web address of NCBI is www.ncbi.nlm.nih.gov.

Smart domain analysis

SMART is a tool used for the prediction of the Domains within a Protein Sequence. It is based upon the similarities within the selected proteins and their conservation upon which the domains are grouped. Web address of smart tool is http://smart.embl-heidelberg.de/.

I-mutant to detect the effect of mutations on the stability

I-Mutant2.0 is a support vector machine (SVM)-based tool for the automatic prediction of protein stability changes upon single point mutations. The predictions are performed starting either from the protein structure or, more importantly, from the protein sequence. The tool would take the protein sequence and the target site as input long with the changed amino acid. The output would be the effect of the specified change on the stability of the protein. Web address of I-mutant tool is http://folding.biofold.org/i-mutant/i-mutant2.0.html.

Modeller

Modeller is a computer program that models three-dimensional structures of proteins and their assemblies by satisfaction of spatial restraints. MODELLER is most frequently used for homology or comparative protein structure modelling. The user provides an alignment of a sequence to be modelled with known related structures and modeller will automatically calculate a model with all non-hydrogen atoms.

Rampage

Structure verification program rampage server (*mordred.bioc.cam.ac.uk/~rapper/rampage.php*) was used to evaluate the 3D-model of PB1 protein. It determines the quality of the predicted structure by assessing various parameters such as lengths, angles and planarity of the peptide bonds, geometry of the hydrogen bonds, and side chain conformations of protein structures as a function of atomic resolution.

RESULTS AND DISCUSSION

The sequence of the PB1 protein has been retrieved from the database and is used for the analysis as hereunder.>gi|8486135|ref|NP_040985.1| polymerase PB1 [*Influenza A* virus (A/Puerto Rico/8/1934(H1N1))]

MDVNPTLLFLKVPAQNAISTTFPYTGDPPYSHGTGTG
YTMDTVNRTHQYSEKARWTTNTETGAPQLNPIDGPLP
EDNEPSGYAQTDCVLEAMAFLEESHPGIFENSCIETM
EVVQQTRVDKLTQGRQTYDWTLNRNQPAATALANTIE
VFRSNGLTANESGRLIDFLKDVMESMKKEEMGITTHF
QRKRRVRDNMTKKMITQRTIGKRKQRLNKRSYLIRAL
TLNTMTKDAERGKLKRRAIATPGMQIRGFVYFVETLAR
SICEKLEQSGLPVGGNEKKAKLANVVRKMMTNSQDT
ELSLTITGDNTKWNENQNPRMFLAMITYMTRNQPEWF
RNVLSIAPIMFSNKMARLGKGYMFESKSMKLRTQIPAE
MLASIDLKYFNDSTRKKIEKIRPLLIEGTASLSPGMMMG
MFNMLSTVLGVSILNLGQKRYTKTTYWWDGLQSSDD
FALIVNAPNHEGIQAGVDRFYRTCKLHGINMSKKKSYI
NRTGTFEFTSFFYRYGFVANFSMELPSFGVSGSNESA
DMSIGVTVIKNNMINNDLGPATAQMALQLFIKDYRYTY
RCHRGDTQIQTRRSFEIKKLWEQTRSKAGLLVSDGGP
NLYNIRNLHIPEVCLKWELMDEDYQGRLCNPLNPFVS
HKEIESMNNAVMMPAHGPAKNMEYDAVATTHSWIPK
RNRSILNTSQRGVLEDEQMYQRCCNLFEKFFPSSSYR
RPVGISSMVEAMVSRARIDARIDFESGRIKKEEFTEIMK
ICSTIEELRRQK

SMART Domain Identification

Using the SMART online tool the domain identification was performed for the PB1 protein. The result is shown below

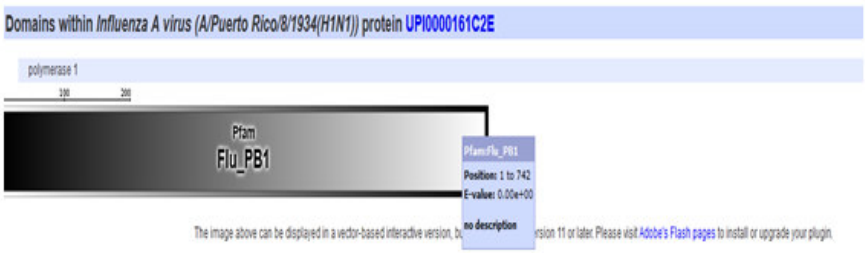

Figure 1
Showing the Domain for the protein as shown by SMART

From the above figure it can be inferred that the protein has only one domain starting from 1 and ending at 742 on the protein sequence. This region also includes the study site 715 Valine, supporting the importance of the site.

I Mutant to detect the effect of mutations on the stability

I mutant tool was used to detect the effect of all the other 19 combinations of amino acids at the 715 position on the stability of the protein. The tool would take the protein sequence and the target site as input long with the changed amino acid. The output would be the effect of the specified change on the stability of the protein.

Position	WT	NEW	Stability	RI	pH	T
715	V	L	Increase	2	7.0	25
715	V	I	Decrease	5	7.0	25
715	V	M	Decrease	6	7.0	25
715	V	F	Decrease	8	7.0	25
715	V	W	Decrease	6	7.0	25
715	V	Y	Decrease	9	7.0	25
715	V	G	Decrease	9	7.0	25
715	V	A	Decrease	7	7.0	25
715	V	P	Decrease	3	7.0	25
715	V	S	Decrease	9	7.0	25
715	V	T	Decrease	9	7.0	25
715	V	C	Decrease	7	7.0	25
715	V	H	Decrease	9	7.0	25
715	V	R	Decrease	7	7.0	25
715	V	K	Decrease	8	7.0	25
715	V	Q	Decrease	9	7.0	25
715	V	E	Decrease	4	7.0	25
715	V	N	Decrease	6	7.0	25
715	V	D	Decrease	6	7.0	25

WT: Aminoacid in Wild-Type Protein
NEW: New Aminoacid after Mutation
RI: Reliability Index
T: Temperature in Celsius degrees
pH: -log[H+]

Figure 2
Showing I Mutant to detect the effect of substitutions on the stability of the protein

Inference

As shown in the above result all the 19 possible substitutions at the 715 position are responsible for a decrease in the stability of the protein. This can conclude that the 715 position is more important according to the stability and functionality and thus an important site for the study.

Modeller

The MODELLER was used for homology or comparative modeling of three-dimensional protein structures. The Alignment of a sequence to be modelled was provided with the known related structures and the modeller automatically calculated a model containing all non-hydrogen atoms. The MODELLER implemented the comparative protein structure modeling by satisfaction of spatial restraints, and could perform many additional tasks, including de-novo modeling of loops in protein structures, optimization of various models of protein structure with respect to a flexibly defined objective function, multiple alignment of protein sequences and/or structures, clustering, searching of sequence databases, comparison of protein structures, etc. The Structure of modelled protein was visualized using Rasmol (Structure visualization tool)

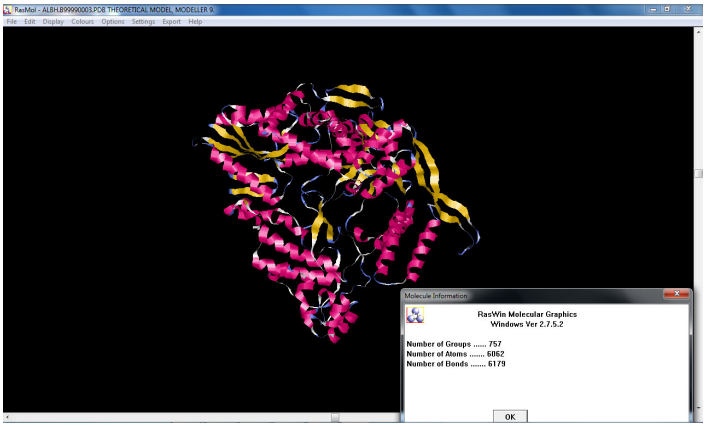

Figure 3
Shows Modeled structure of PB1

Validation of the generated model

RAMPAGE is a program for visualising and assessing the Ramachandran plot of a protein structure. On the basis of a manually curated set of high-quality protein structures (from the Richardson's Group at Duke University) and a number of filters (such as B-factor cutoff and van der Waals clashes), reference phi/psi plots were derived for Gly, Pro, pre-Pro and general (other) residue types, and subdivided into "favoured", "allowed" and "outlier" regions. Residues in the uploaded PDB file that fall into the "allowed" and "outlier" regions are listed, and a picture of the Ramachandran plot is displayed. The output can also be printed as a high-resolution multi-colour Adobe PDF or PostScript file, containing the general plot and the four separate plots (2 pages). Furthermore, the validation studies of the generated model of PB1 using RAMPAGE server showed 97.1 % residues in most favoured regions, 2.6 % in allowed regions and 0.3 % of amino acid residues were found in the disallowed region. Most of the residues are present in favoured region. This is considered as good model.

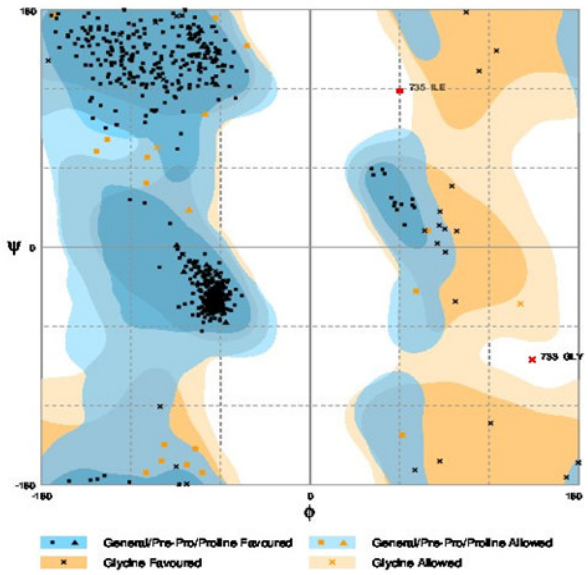

Figure 4
Shows Ramachandran plot of modeled structure

CONCLUSION

Due to the reoccurrence of H1N1 pandemic there is a need to study the evolution of different H1N1 proteins so as to make out spread of the disease. Influenza A virus is a lipid-enveloped orthomyxovirus and a cause of human disease by strains that arise through seasonal variation and through pandemic infection resulting from viral adaptation that introduces new influenza viruses into the human population[4].The current study deals with the insilico sequence and structure analysis of PB1 protein of *Influenza A* virus by various tools. This protein is the subunit of polymerase protein. It is involved in both transcription and replication of the genome RNA. Based on the finding, it could be concluded that the protein was mutated and 3D structure was predicted. The predicted structure can be used to know more about interaction of PB1 protein domains with ligands in future.

FUNDING/ACKNOWLEDGEMENT

Financial support to Dr. K.Akila, Principal Investigator, Minor Research project (MRP-5112/14 [SERO/ UGC]), University Grant Commission (UGC), New Delhi, India is acknowledged. The author is also grateful to the Department of Biotechnology and the Management of Bishop Heber College, Trichy for their support.

CONFLICT OF INTEREST

Conflict of interest declared none.

REFERENCES

1. Akila K, Balamurugan P, Rajasekaran E. The nature of proteins in influenza.2012 0ct 25;4(10A):991-94.
2. Banerjee R, Roy A, Das S, Basak S. Similarity of currently circulating H1N1 virus with the 2009 pandemic clone: Viability of an imminent pandemic. Infection, Genetics and Evolution. 2015 Jun 30;32:107-12.
3. Binh NT, Wakai C, Kawaguchi A, Nagata K. The N-terminal region of influenza virus polymerase PB1 adjacent to the PA binding site is involved in replication but not transcription of the viral genome. Frontiers in microbiology. 2013 Dec 18;4:398.
4. Calder LJ, Wasilewski S, Berriman JA, Rosenthal PB. Structural organization of a filamentous influenza A virus. Proceedings of the National Academy of Sciences. 2010 Jun 8;107(23):10685-90.
5. Clancy S. Genetics of the influenza virus. Nature Education. 2008;1(1):83.
6. Dawood FS, Iuliano AD, Reed C, Meltzer MI, Shay DK, Cheng PY, Bandaranayake D, Breiman RF, Brooks WA, Buchy P, Feikin DR. Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study. The Lancet infectious diseases. 2012 Sep 30;12(9):687-95.
7. Kawaoka Y, Krauss S, Webster RG. Avian-to-human transmission of the PB1 gene of influenza A viruses in the 1957 and 1968 pandemics. Journal of virology. 1989 Nov 1;63(11):4603-8.
8. Liu Y, Zhang L, Zhou Y. Comparative Analysis of H1N1 Avian Influenza Virus by Multiple Sequence Alignment and Support Vector Machine. J Gene Ther. 2014;2(1):4.
9. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM. Stereochemical quality of protein structure coordinates. Proteins: Structure, Function, and Bioinformatics. 1992 Apr 1;12(4):345-64.
10. Noton SL, Medcalf E, Fisher D, Mullin AE, Elton D, Digard P. Identification of the domains of the influenza A virus M1 matrix protein required for NP binding, oligomerization and incorporation into virions. Journal of General Virology. 2007 Aug 1;88(8):2280-90.
11. Resa-Infante P, Jorba N, Coloma R, Ortín J. The influenza virus RNA synthesis machine: advances in its structure and function. RNA biology. 2011 Mar 1;8(2):207-15.
12. Smith GJ, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, Ma SK, Cheung CL, Raghwani J, Bhatt S, Peiris JM. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature. 2009 Jun 25;459(7250):1122-5.
13. Steitz TA. Structural biology: A mechanism for all polymerases. Nature. 1998 Jan 15;391(6664):231-2.
14. Taubenberger JK, Reid AH, Lourens RM, Wang R, Jin G, Fanning TG. Characterization of the 1918 influenza virus polymerase genes. Nature. 2005 Oct 6;437(7060):889-93.
15. Xiang Z. Advances in homology protein structure modeling. Current Protein and Peptide Science. 2006 Jun 1;7(3):217-27.
16. Xu J, Zhong HA, Madrahimov A, Helikar T, Lu G. Molecular phylogeny and evolutionary dynamics of influenza A nonstructural (NS) gene. Infection, Genetics and Evolution. 2014 Mar 31;22:192-200.

Reviewers of this article

Dr.R.Jasmine

Asst Professor in Zoology
Bishop Heber College
Trichy-17

Mr. Anubrata Paul M.Sc. Biotech (Research)

Department of Biotechnology, Natural Products Research Laboratory, Centre for Drug Design Discovery & Development (C-4D) , SRM University, Delhi-NCR,Sonepat.

Prof.Dr.K.Suriaprabha

Asst. Editor , International Journal of Pharma and Bio sciences.

Prof.P.Muthuprasanna

Managing Editor , International Journal of Pharma and Bio sciences.

We sincerely thank the above reviewers for peer reviewing the manuscript