British Biotechnology Journal
3(4): 545-555, 2013

SCIENCEDOMAIN *international*
www.sciencedomain.org

Structural Comparison of Wild and Mutated Matrix Protein of Nipah Virus via *in-silico* Approach

Sudhir kumar Pal[1*]

[1]*Department of Biotechnology and Bioinformatics, Padmashree Dr. D. Y. Patil Vidyapeeth, Plot no 50, Sector 15, CBD Belapur, Navi Mumbai -400 614, Maharashtra, India.*

Author's contribution

The only author performed the whole research work. Author SKP wrote the first draft of the paper. Author SKP read and approved the final manuscript.

Research Article

Received 25th March 2013
Accepted 14th July 2013
Published 15th August 2013

ABSTRACT

Aims: To make three dimensional structure of native and mutated matrix protein (M) of Nipah Virus (NiV) and to establish conformational and functional comparison between the two.
Study Design: All *in-silico* analysis were performed using various online and offline software.
Place and Duration of Study: Department of Biotechnology and Bioinformatics, Padmashree Dr. D.Y.Patil Vidyapeeth, Navi Mumbai, February 2013.
Methodology: The protein's physicochemical properties were characterized and 3D model of both the normal and the mutated protein were created using Phyre2. Single point mutation of S147G is recorded which leads to altered structure formation. Both the models were evaluated and compared conformationally.
Results: 4G1GB based structures were modeled by phyre2 and minimized energies recorded were-16760.041 kJ/mol for native and -16563.029 kJ/mol for the mutated protein. Structure validation proved that both the native and mutated structures were reliable. Formation of 3 H-bonds make mutated M structure slightly more stable than the native one.
Conclusion: NiV, one of deadliest pathogen, needs to be checked immediately. More information gain is needed by performing wet lab analysis. This work might help understand the functional difference between native and mutated M protein and can be

**Corresponding author: Email: leo.oberoi@gmail.com;*

used as the potent drug target via applying rational drug designing approach.

Keywords: Nipah virus; matrix protein; PHYRE2.

1. INTRODUCTION

Nipah (NiV) virus, named after the place where it first emerged in Malaysia in 1998 [1], is a member of Paramyxoviridae family. After Malaysia outbreak with 106 killings, NiV has caused various outbreaks in Bangladesh and Hindustan from 2001-2008 showing 67-92% fatality rate [2,3,4,5,6]. Transmission of NiV to pigs and then humans is carried out by their natural reservoirs, bats belonging to the genus *Pteropus* containing several species like flying fox and fruit bats [5]. As NiV infect both humans and other domestic animals including dogs and cats, it seems to be a threat to both public health and agriculture [6].

NiV infected patients display rapid onset of symptoms within 7-10 days of infection. The most common are aseptic meningitis and encephalitis with fever, headache and vomiting, and reduced level of consciousness. NiV infected patients has characteristic brain magnetic resonance imaging (MRI) scan findings and CSF pleocytosis with elevated protein levels, the multifocal lesions in the brain and segmental myoclonus appear unique to NiV infection [7,8].

Paramyxoviruses replicate in the cytoplasm of infected cells, and newly produced virions are released from the plasma membrene [9]. The viral matrix protein (M) is considered to play a critical role in paramyxovirus assembly and appears to be the driving force for virion budding [10]. The co-expression of Fusion (F) and attachment (G) proteins along with M revealed a shift in their distribution across the gradient, indicating association with M in VLPs. NiV M, F, and G each possess some ability to bud from expressing cells, and that co-expression of these viral proteins results in a more organized budding process with M playing a central role [11].

With the so called important role, Matrix protein could be a vital drug target. But unfortunately the crystal 3-D structure of M is not yet discovered. Some researchers try to make the structure with *in-silico* approaches. In this paper we we have tried to build a three dimensional structure of M protein with some new and improved *in-silico* approaches.

2. METHODOLOGY

The amino acid sequence for Matrix protein of Nipah virus (Uniprot-ID: Q9IK90) was downloaded from Uniprot [12] in the FASTA format. Given below is the said sequence.

>sp|Q9IK90|MATRX_NIPAV Matrix protein OS=Nipah virus GN=M PE=3 SV=1

```
MEPDIKSISS   ESMEGVSDFS   PSSWEHGGYL   DKVEPEIDEN   GSMIPKYKIY   TPGANERKYN   NYMYLICYGF
VEDVERTPET   GKRKKIRTIA   AYPLGVGKSA   SHPQDLLEEL   CSLKVTVRRT   AGSTEKIVFG   SSGPLNHLVP
WKKVLTSGSI   FNAVKVCRNV   DQIQLDKHQA   LRIFFLSITK   LNDSGIYMIP   RTMLEFRRNN   AIAFNLLVYL
KIDADLSKMG   IQGSLDKDGF   KVASFMLHLG   NFVRRAGKYY   SVDYCRRKID   RMKLQFSLGS   IGGLSLHIKI
NGVISKRLFA  QMGFQKNLCF   SLMDINPWLN   RLTWNNSCEI  SRVAAVLQPS   IPREFMIYDD  VFIDNTGRIL
KG
```

Physicochemical characteristics were analyzed by Expasy's online 'protparam' tool [13]. The parameters computed by protparam included the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half life, instability index,

aliphatic index, and grand average of hydropathicity (GRAVY). The grand average of hydropathy (GRAVY) value for a peptide or protein is calculated as the sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence. Negetive values of GRAVY indicate the hydrophilic nature of protein [14]. More is the negative value higher is the hydrophilicity.

The three dimensional structure of M was built using the protein homology/analogy recognition engine 2 (phyre2) [15]. While running intensive mode, phyre2 selected Newcastle disease virus matrix protein chain B (PDB ID: 4g1g at 2.2 Å) as template to model M based on heuristics to maximise confidence, percentage identity and alignment coverage. 32 residues were modelled by *ab initio* modelling which seems to be unreliable. This modelled 3-D structure was then subjected to FoldX [16] to provide a fast and quantitative estimation of the importance of the interactions contributing to the stability of proteins. FoldX uses a full atomic description of the structure of the proteins. The different energy terms, taken into account in it, have been weighted using empirical data based on protein engineering experiments.

Cleaned/Repaired protein structures were then minimized energetically using Swiss-PDBViewer [17]. The energy of the minimized protein was then recorded. The repaired and energy minimized proteins were cross checked using Ramachandran plot (RAMPAGE) [18] and ProSA [19]. The Z-score calculated by ProSA web server indicates overall model quality and measures the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations. *Z*-scores outside a range characteristic for native proteins indicate erroneous structures [19].

Different drug binding sites are found by Q-site finder [20] which provides us the information about the potent ligand binding sites on the surface of the protein. It works by binding hydrophobic (CH3) probes to the protein, and finding clusters of probes with the most favourable binding energy. These clusters are placed in rank order of the likelihood of being a binding site according to the sum total binding energies for each cluster.

Knowledge about the natural variant strain-isolate NiV/MY/99/VRI-0626 of M which is having an S147G mutation is obtained from Uniprot. Same above methodology has been followed to get the information regarding structural and functional variability in the mutated protein occurred due to the mutation. All the results are recorded in the form of tables and figures discussed in the results and discussion part of this paper.

3. RESULTS AND DISCUSSION

The results obtained by the study of physicochemical parameters of both the native and mutated M protein obtained by protparam online tool [21] are shown Table 1. The instability index shows a slight variation in both the structures. We can easily see the increase in instability of the mutated M protein. Although the increment is not that much significant structurally but it might affect protein functionality at a higher level as demonstrated by the RMS calculation (Fig. 1). The RMS deviation value of wild and mutated protein is calculated by spdbv and obtained to be 9.88 Å (nb atoms involved 1408). This shows a larger variation in the structural organisation of amino acids leading to functional deviation/instability of the mutated protein.

Table 1. Physicochemical characterization through Expasy's Protparam tool [21]

Property	Value	
	Wild	**mutated**
Number of amino acids	352	352
Molecular weight	39928.2	39898.2
Theoretical pI	9.31	9.31
Total number of negatively charged residues (Asp+Glu)	36	36
Total number of positively residues (Arg+Lys)	48	48
Extinction coefficient assuming all pair of Cys form Cysteines	43235	43235
Extinction coefficient* assuming all Cysteine residues are reduced	42860	42860
Instability index	29.53	29.64
Aliphatic index	90.26	90.26
Grand average of hydropathicity	-0.211	-0.210

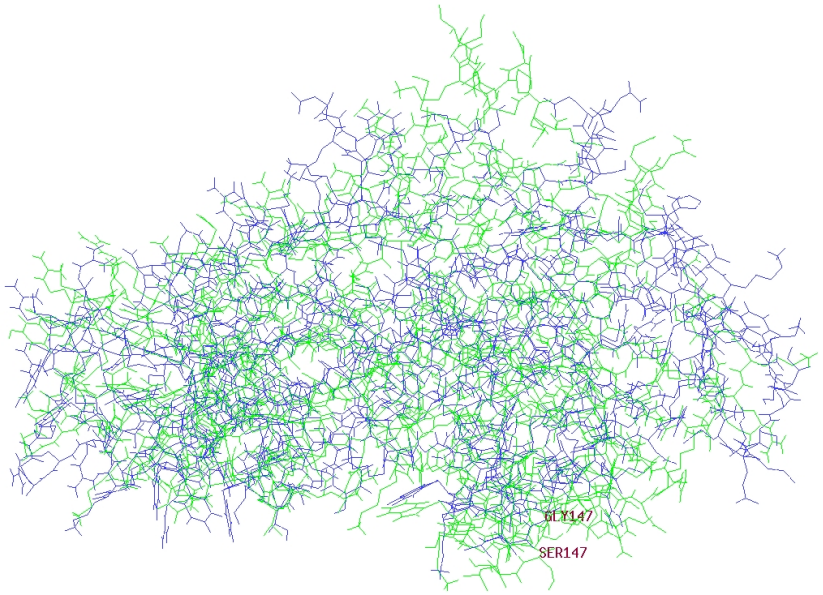

Fig. 1. RMS deviation analysis by superimposing native (green) and mutated (blue) matrix protein. Deviation in the orientation of amino acids involved in mutation

According to the protparam analysis (Fig. 2), an insignificant decrease in negative value of GRAVY in the case of S147G mutation is obtained which points to a slightly increased hydrophobicity of the protein. This change in hydrophobicity at par level might result in the disturbed residue profiling. A protein is said to be stable if it is having instability index smaller than 40 [21,22] and both the native and mutated variants are having lower values indicating stable conformations. High aliphatic index (90.26) suggests that the proteins may be stable for a wide temperature range [21].

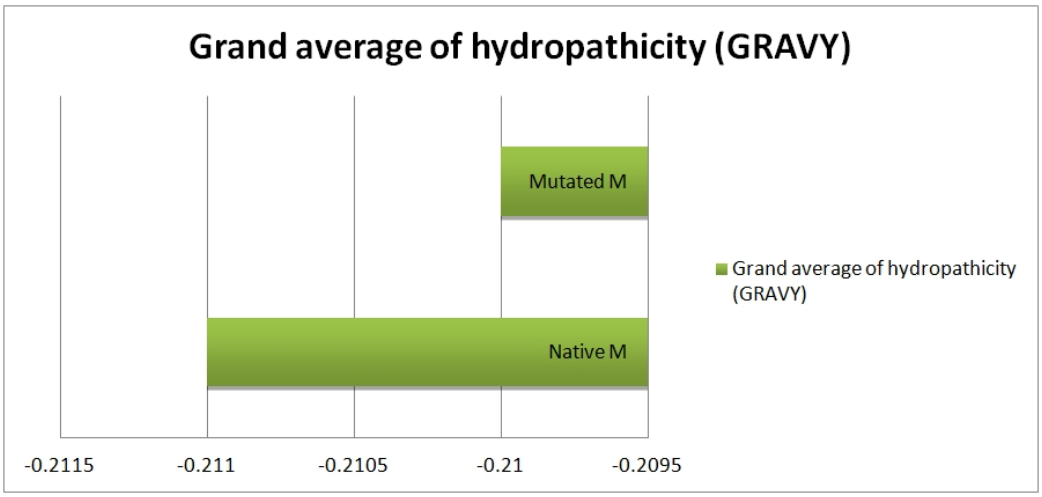

Fig. 2. hydropathy of native and mutated matrix of nipah virus

The homology modelling of both the variants of M was performed by PHYRE 2 server (Figs. 3,4). The visualizar used in these figures was Accelrys Discovery Studio client 3.5 [23]. Energy obtained after minimization of modelled 3-D structure was -16760.041 kJ/mol for native and -16563.029 kJ/mol for the mutated protein. The evaluation of the structures and their comparative analysis for both the WT and MT M protein was performed using Ramachandran plot (RAMPAGE) (Fig. 5).

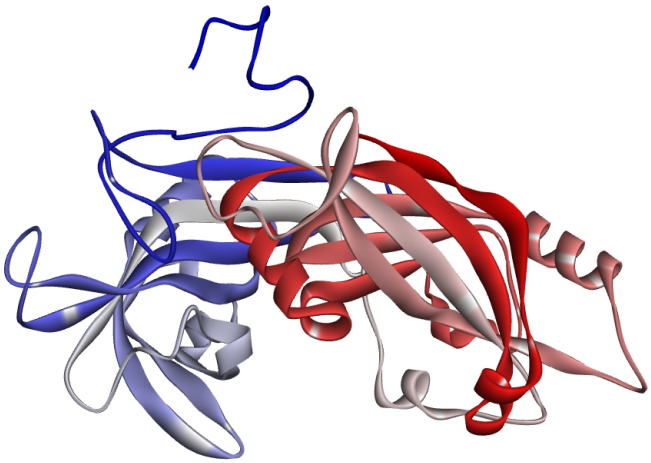

Fig. 3. Three dimensional structure of Native Matrix protein of Nipah Virus modelled using phyre2 visualized in discovery studio visualizer [23]. Colouring is according to the N to C-terminal from blue to red

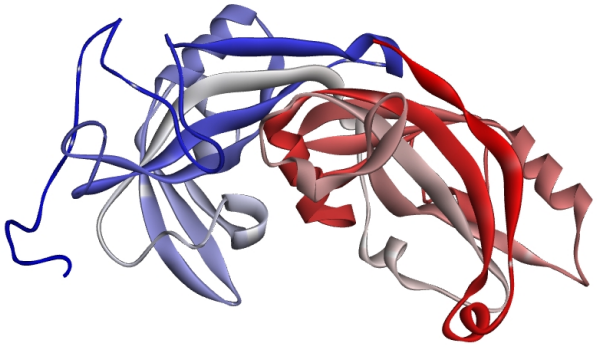

Fig. 4. Three dimensional structure of mutated M of NiV modelled using phyre2 visualized in discovery studio visualizer [23]. Colouring is according to the N to C-terminal from blue to red

Around 310 (88.6%) residues in favoured region, 25 (7.1%) residues in allowed region, 15 (4.3%) residues in outlier region were calculated by Rampage for WT M. Around 315 (90.0%) residues in favoured region, 23 (6.6%) residues in allowed region, 12 (3.4%) residues in outlier region were calculated by Rampage for MT M (Fig. 6). Red squares indicate the residues lies in outlier region (Fig. 5). This analysis shows that mutated M protein structure is more stable than the structure of native protein. Serine (S), which has a polar uncharged side chain, make two Hydrogen bonds in the native protein and Glycine (G), which has hydrogen as its side chain, forms three hydrogen bonds with the surrounding residues (Fig. 7). Difference in the orientation of Ser-147 also account for the less number of Hydrogen bonds formation as clear in the Fig. 7 Thus leads to the generation of a pull down force which might flip the chain as shown in Figs. 3 and 4.

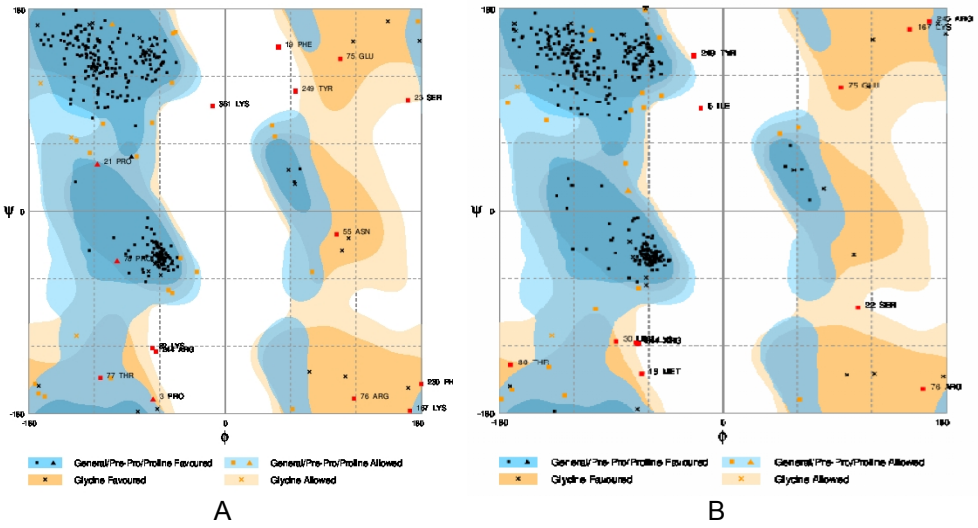

Fig. 5. Ramachandran plot of native (A) and mutated Matrix protein (B) generated by RAMPAGE

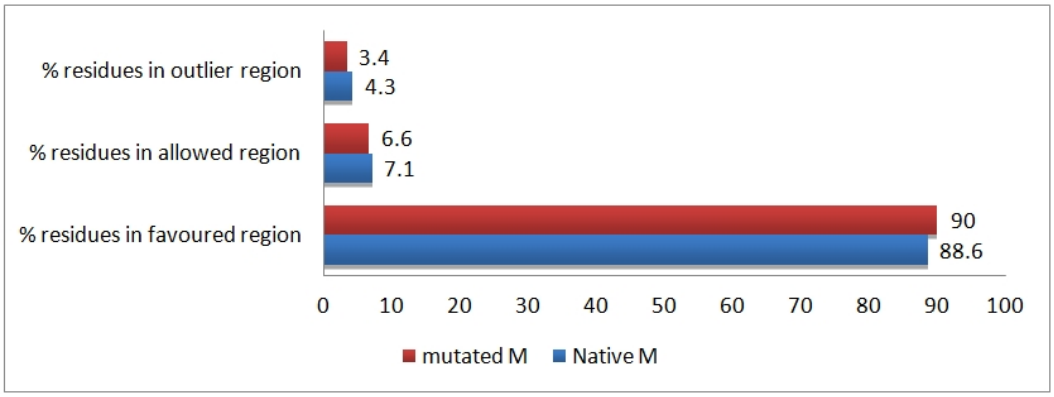

Fig. 6. Results of residue profiling of native and mutated Matrix protein calculated by RAMPAGE

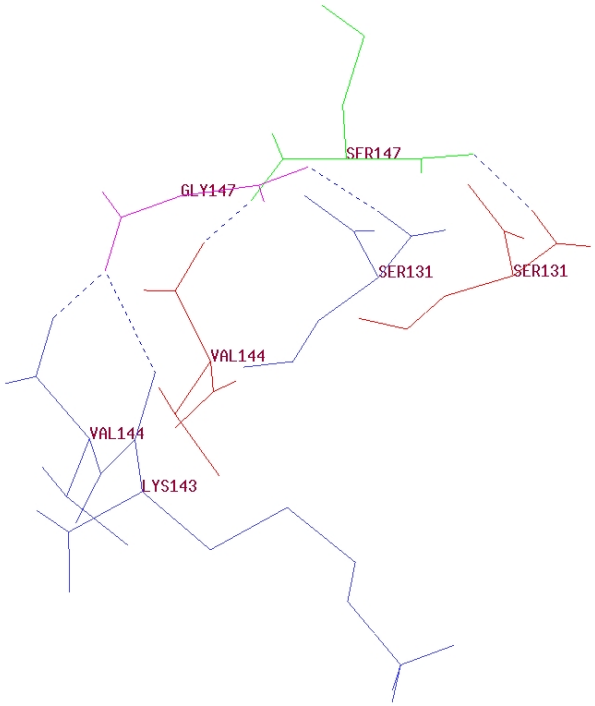

Fig. 7. Hydrogen bonding pattern by SPDBV software. In this figure blue layer is mutated and red is native Matrix protein. G-147 (in pink) shows three H-bonds (dashed blue lines) and S-147 (in green) shows two H-bonds formation with neighbouring residues

The Z-score calculated by ProSA web server was -6.41 for WT protein and -6.55 for MT. The Z-score indicates that overall model quality is reliable. The values are displayed by the shown plots that contain the *z*-scores of all experimentally determined protein chains in

current PDB (Fig. 8). The black spots shown in these plots give us information that both the proteins are in the optimum ranges. Fig. 9 shows local model quality by plotting energies as a function of amino acid sequence positions. The positive values correspond to problematic or erroneous parts of the input structure. A plot of energies of single residue usually contains large fluctuations and is of limited value for model evaluation. Therefore the plot is smoothed by calculating the average energy over each 40-residue fragments, which is then assigned to the 'central' residue of the fragment at position *i+19.* A second line with a smaller window size of 10 residues is shown in the background of the plot [19].

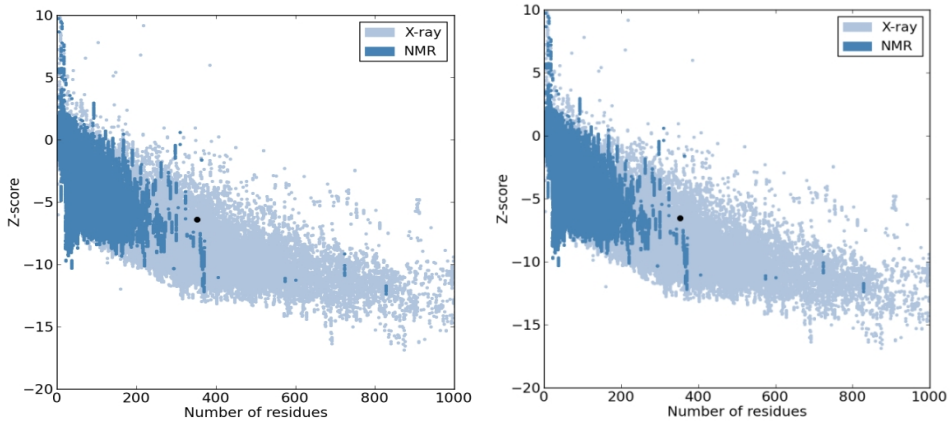

Fig. 8. Z-score plot of native (A) and mutated Matrix protein (B) of NiV generated through ProSA web-server. Black spots are the structured models of matrix protein

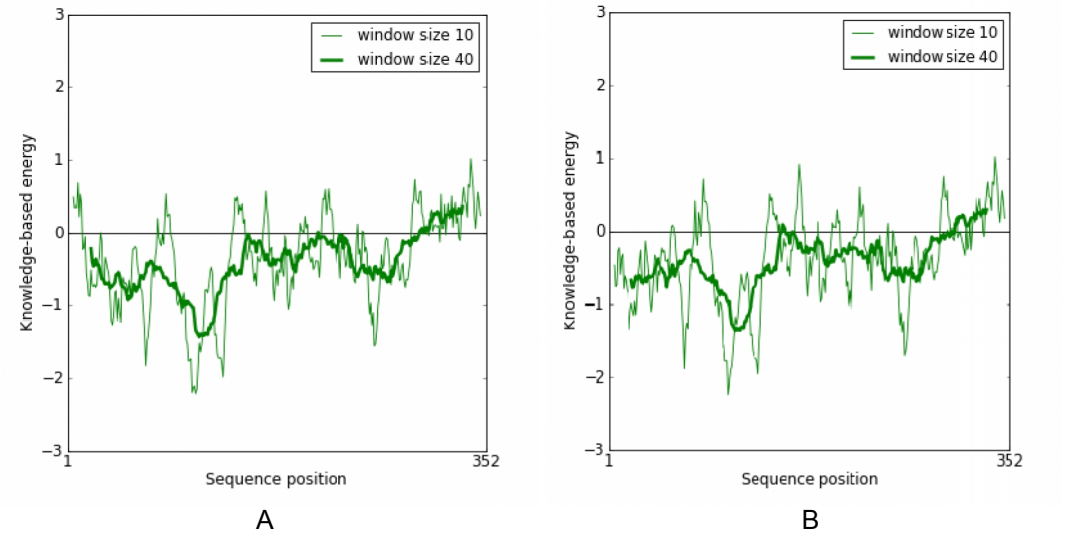

Fig. 9. Plot of energies as a function of amino acid sequence position for native (A) and mutated Matrix protein (B) generated by ProSA web-server

Q-site finder found 10 ligand binding sites (Table 2). Site number 1 (Site Volume: 635 Cubic Angstroms) with respect of whole protein (Protein Volume: 32683 Cubic Angstroms) was the largest binding site and Site number 10 (Site Volume: 157 Cubic Angstroms) the smallest. On the basis of literature survey it is known that residue number 147 where S147G mutation takes place lies in none of the obtained binding sites. This suggests that the mutation might have an allosteric effect on the protein structure and drug binding regions.

Table 2. Information of binding site obtained through Q-site finder

Site information	Native M protein	Mutated M protein
Total number of binding sites	10	10
Total protein volume (cubic Å)	32683	32383
Largest site (cubic Å)	635	383
Smallest site (cubic Å)	157	172

4. CONCLUSION

Recently, Nipah virus has emerged as one of the deadliest viruses due to globalization and human encroachment into native wild-life habitats [24]. Henipa-viruses are classified as biosafety level 4 (BSL4) pathogens due to their high virulence and absence of therapeutics or vaccines. NiV is classified as a Category C priority pathogen by the NIAID Biodefense Research Agenda for its bio- and agro-terrorism potential [24, 25]. Therefore, a great need of hour is to develop efficient therapeutics to take over these viruses and to stop their deadly game. This research might help in doing so.

ACKNOWLEDGEMENT

The author would like to thank his parents, Dr. Vijay Nema and friend Divya Pandey for their support, guidance and ideas, and to all other people for their important feedback. I thank all the reviewer of Sciencedomain team to review and correct this manuscript.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Vigant F, Lee B. Hendra and Nipah infection: pathology, models and potential therapies. Infect Disord Drug Targets. 2011 June 1;11(3):315–36.
2. Chua, KB, Bellini WJ, Rota PA, Harcourt BH, Tamin A, Lam SK et al. Nipah virus: a recently emergent deadly paramyxovirus. Science 2000;288:1432–35.
3. Wild TF. Henipaviruses: a new family of emerging Paramyxoviruses. Pathol Biol (Paris). 2009;2:188–96. [PubMed: 18511217].
4. Wolf MC, Negrete OA, Lee B. Pathobiology of henipavirus entry: insights into therapeutic strategies. Future Virol. 2007;3:267–82.
5. Chua, KB, Koh CL, Hooi PS, Wee KF, Khong JH, Chua BH et al. Isolation of Nipah virus from Malaysian Island flying-foxes. Microbes Infect. 2002;4:145–51.

6. Ciancanelli MJ, Basler CF. Mutation of YMYL in the Nipah virus matrix protein abrogates budding and alters subcellular localization. J Virol. 2006;80:12070–78;doi:10.1128/JVI.01743-06.
7. Chua KB, Lam SK, Tan CT, Hooi PS, Goh KJ, Chew NK et al. High mortality in Nipah encephalitis is associated with presence of virus in cerebrospinal fluid. Ann Neurol. 2000;5:802–5. [PubMed: 11079547]
8. Goh KJ, Tan CT, Chew NK, Tan PS, Kamarulzaman A, Sarji SA et al. Clinical features of Nipah virus encephalitis among pig farmers in Malaysia. N Engl J Med. 2000; 17:1229–35. [PubMed: 10781618].
9. Lamb RA, Kolakofsky D. Paramyxoviridae: the viruses and their replication. In: Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Straus SE, editors. Fields Virology. 4th ed. Philadelphia, PA, Lippincott Williams & Wilkins; 2001:1305-40.
10. Takimoto T, Portner A. Molecular mechanism of paramyxovirus budding. Virus Res. 2004;106:133–45.
11. Patch JR, Crameri G, Wang LF, Eaton BT and Broder CC. Quantitative analysis of Nipah virus proteins released as virus-like particles reveals central role for the matrix protein. Virol. J. 2007;4:1.
12. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S et al UniProt: the Universal Protein knowledgebase. Nucl Acids Res. 2004;32:D115-19.
13. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD et al. Protein identification and analysis tools in the ExPASy server. Meth Mol Biol. 1999;112:531-52.
14. Barik BP, Tayung K. Molecular differentiation of *Fusarium* spp. with varied lifestyles based on TEF 1 alpha gene sequence analysis. Interdisciplinary sciences, computational life sciences. 2012;4:201-8.
15. Lawrence AK, Michael J E. Protein structure prediction on the Web: a case study using the Phyre server. Published online 26 February 2009;doi:10.1038/nprot.2009.2.
16. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. Nucl. Acids Res. 2005;33:W382-8.
17. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis. 1997;18:2714-23.
18. Lovell SC, Davis IW, Arendall WB III, de Bakker PIW, Word JM, Prisant MG et al. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins: Structure, Function & Genetics. 2002;50:437-50.
19. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucl acids Res. 2007;35: W407-10.
20. Laurie AT, Jackson RM. Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. Bioinformatics. 2005;21:1908-16.
21. Gupta S, Kumar S, Saxena S, Yadav R, Thakur A, Verma OP, et al. Insilico analysis and 3D structure prediction of matrix protein of Nipah virus. J. Adv Bioinform Appl Res.;ISSN 0976-2604.Online ISSN 2278 – 6007. 2013;4(1):391-5.
22. Guruprasad K, Reddy BVP, Pandit MW. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Prot Eng. 1990;4:155-64.
23. Ref. Accelrys Software Inc., Discovery Studio Modelling Environment, Release 3.5, San Diego: Accelrys Software Inc.; 2012.

24. Aguilar HC, Lee B. Emerging paramyxoviruses: molecular mechanisms and antiviral strategies. Expert Rev Mol Med. 2012;13:e6. doi:10.1017/S1462399410001754.
25. Lam SK. Nipah virus-a potential agent of bioterrorism? Antiviral Res. 2003;57(1–2):113–9.[PubMed: 12615307].

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history.php?iid=243&id=11&aid=1867