ISSN 0975-6299 Vol 2/Issue 1/ Jan-Mar 2011

International Journal of Pharma and Bio Sciences

RESEARCH ARTICLE

BIOINFORMATICS

COMPARATIVE COMPUTATIONAL STUDIES AND *INSILICO* IDENTIFICATION OF POLYPROTEIN IN JAPANESE ENCEPHALITIS VIRUS

Corresponding Author

ABHISHEK KR. MALAKAR

National bureau of fish genetic resources, lucknow, uttar pradesh, india.

Co Authors

MANISHA MISHRA[1], ASHISH KR. JAISWAL[3] AND PALLAVI[2]

[1]amity university, lucknow, uttar pradesh, india
[2]national bureau of fish genetic resources, lucknow, uttar pradesh, india.
[3]national jalma institute for leprosy & other mycobacterial diseases –agra – 282001- india.

ABSTRACT

Japanese encephalitis virus (JEV) is the major form of viral encephalitis in much of the South-East Asia, India and China. The disease is caused by a mosquito-borne virus known as Japanese encephalitis virus (JEV). The mortality rate in Japanese encephalitis is 30% while 10 - 15% of patients make a full recovery. Japanese encephalitis virus contains a single positive-strand RNA genome nearly 11 kb in length and is not formally thought to generate subgenomic RNA molecules during replication. The envelope (E) polyprotein of JEV interacts with a cellular receptor and mediates membrane fusion to allow viral entry into target cells, thus eliciting neutralizing antibody response. Despite its importance, the three dimensional (3D) structure of polyprotein has not yet been reported.

Comparative studies are useful in the prediction of validated 3D structure of a query protein (C5NU03). For the modeling, template protein was obtained by Protein Data Bank (PDB) database, template protein pdb|3I50|chain A having identity (76%) and E value (0.0). From Ramachandran plot analysis it was found that the portion of residues falling into the most favoured regions was (73.0%). These results indicate that the predicted and validated structure (after comparative study) is useful in structure based drug designing, protein-DNA interactions, study of protein-protein interactions, protein-ligand binding and analyzed using other techniques.

ISSN 0975-6299 Vol 2/Issue 1/ Jan-Mar 2011

KEYWORDS

Japanese encephalitis virus (JEV), Polyprotein, Homology modeling, MODELLER & Ramachandran plot.

INTRODUCTION

Japanese encephalitis virus (JEV) is an arthropod-borne flavivirus and is a causative agent of central nervous diseases such as meningitis and severe encephalitis[1]. JEV belongs to the genus Flavivirus of the family Flaviviridae. JEV is a mosquito borne neurotropic flavivirus, endemic-epidemic throughout Asia[2]. Severe JEV infection is marked by quick onset, headache, high fever, neck stiffness, occasional convulsions (especially in infants) and spastic (but rarely flaccid) paralysis, aseptic meningitis, encephalitis and even death[3]. It is transmitted by female mosquitoes of genera Culex, Anopheles, Aedes monsonia and Armegeres mosquitoes. JE virus infection is wide spread in southern states of India viz. Andhra Pradesh, Tamil Nadu, Uttar Pradesh and Karnataka. This disease is very dangerous and causes inflammation of the brain (encephalitis). This swelling can lead to death or brain damage. There is a 30% mortality rate for people who get sick from Japanese Encephalitis, and only 10-15% of patients make a full recovery. However, only 1 in every 30-300 people bitten by an infected mosquito becomes sick while infection rates are small, of those who develop symptoms the disease is fatal in around 25% usually occurring within the first 10 days. Of those who survive up to 50% are left with brain damage. More than 35,000 cases and 10,000 deaths are reported annually from the region but official reports undoubtedly underestimate the true number of cases[4].

Japanese encephalitis virus contains a plus-sense, single-stranded RNA genome of about 11 kilobases[5,6]. The genomic RNA contains a single open reading frame capable of encoding a polyprotein of about 3400 amino acids that is subsequently cleaved, coand post-translationally, by both host and viral proteases, into several structural and nonstructural viral proteins. A lot of information about JEV proteins is derived from studies on other flaviviruses and is assumed to apply to JEV because of the high level of similarity of its genome structure and organization to other flaviviruses[7,8]. Thus, the JEV polyprotein is cleaved to produce three structural (capsid, C; membrane, M; and envelope, E) and seven nonstructural (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) proteins. The assembly of JEV in the endoplasmic reticulum is followed by modification of the two polyproteins E & prM and virion export through the secretory pathway. The E protein plays a major role in virus assembly, adhesion, receptor binding and membrane fusion, hemagglutination inhibition (HI) and induction of neutralizing antibodies[9,10,11]. Therefore, the E protein is the principal target of neutralization by specific antibodies against JEV infection[12,13]. E proteins of JEV expressed in different viral vector systems such as vaccinia virus, sindbis virus, and baculovirus have elicited high levels of neutralizing antibodies against JEV infection and have been tested as second generation JEV vaccines in mice[14,15,16]. Hence, E protein of JEV apparently plays a crucial role in viral pathogenesis by determining the cellular susceptibility and organ tropism of the virus[17,18].

Bioinformatics uses computational approach to answer the biological problems. Answering these questions requires that investigators take advantage of large, complex

data sets (both public and private) in a rigorous fashion to reach valid, biological conclusions[19,20,21]. With the explosion of sequence and structural information available to researchers, the field of bioinformatics is playing an increasingly large role in the study of fundamental biomedical/biotechnological problems. Homology modeling is basically used for the prediction of protein structure and it constructs an atomic-resolution model of a protein from amino acid of query sequence. The quality of the homology model is dependent on the quality of the sequence alignment and template structure[22].

Since there was no structure reported to this protein, the main aim of this study is to predict validated three dimensional structures from the polyprotein sequence by comparative studies. Modeling tools and related structural data available on the online databases are used for the comparative studies between five structures.

MATERIALS & METHOD

The amino acid sequence of the polyprotein of Japanese encephalitis virus was obtained from the protein database of Protein Information Resource (PIR) http://pir.georgetown.edu (Acc. No: C5NU03_9FLAV). It was ascertained that the three-dimensional structure of the protein was not available in Protein Data Bank[20]; hence the present exercise of developing the 3D model of the polyprotein of JEV was undertaken. Important properties of polyprotein were calculated using PROTPARAM[23]. The protein is 517 amino acids in length.

An attempt was made to find a suitable template protein for the modeling of the target protein. The template protein was searched through Protein Data Bank (PDB), which is an online database for searching similar sequences, based on sequence and structure-wise similarity. From the homology searching, template was selected. The X-ray structure of the Crystal structure of the West Nile Virus envelope glycoprotein in complex with the E53 antibody Fab (PDB entry: 3I50) was selected as template proteins[24].

MODELER is a computer program used in producing homology models of protein tertiary structures and sometimes in quaternary structures[25,22,26]. Modeling is a method of designing 3D model for a protein of unknown structure based on one or more related protein of known structure. MODELER is a computer program that most frequently used for homology modeling. Homology modeling requires at least one sequence of known 3D structure with significant similarity with the target sequence. MODELER is a command-line tool and has no graphical user interface. It requires a script file containing MODELER commands. This is an ordinary Python 2.5 script. The script file contains commands for MODELER. A script files to produce models of target sequence from the known structure. Target sequence is used as a query and selection of target is according to the lowest resolution value is done by running the "build_profile.py" file in MODELER. From the log file of "build_profile.py" the sequence having lowest e-value and alignment score greater than 45% considered as the templates and selected templates are compared with the query in "compare.py" file by running the modeler (Table 1). After this, template having the lowest resolution value selected as a template and finally "model-single.py" file is running in modeler for models generation, best model is selected according to its highest molpdf & lowest Dope (Discrete optimized potential energy) score (Table 2).

Table 1
List of templates and related information's are shown below.

S. No.	Templates	No. of Residues in Templates	Resolution	E value	Percentage Similarity
1.	3I50	402	3.00	0.0	76
2.	2I69	403	3.11	0.0	76
3.	2HG0	394	3.00	0.0	76
4.	2OF6	400	-	0.0	76
5.	3IXX	400	-	0.0	76

Table 2
Dope score of models are listed below.

S. No.	Models generated by modeler	molpdf	Dope score
1.	C5NU03.B99990001.pdb	13219.04980	-37003.92188
2.	C5NU03.B99990002.pdb	14077.04785	-35211.12891
3.	C5NU03.B99990003.pdb	13700.92969	-35959.30469
4.	C5NU03.B99990004.pdb	14582.18652	-35016.59375
5.	C5NU03.B99990005.pdb	14899.74121	-33954.00000

The constructed model of polyprotein of JE virus was examined for validation using different criteria. Model evaluation is done by using with VERIFY3D[27] and Ramachandran plot at PROCHECK[21]. Backbone conformation was evaluated by the inspection of the Psi/Phi Ramachandran plot obtained from PROCHECK analysis. The Swiss-PdbViewer energy minimization test was applied to check for energy criteria in comparison with the potential of mean force derived from a large set of known protein structures. Packing quality of the refined structure was investigated by the calculation of PROCHECK Quality Control value. The Ramachandran plot of phi/psi distribution in the model is developed using PROCHECK for checking non-GLY residues at the disallowed regions. Standard bond lengths and bond angles of the model were determined using WHATIF[28].

RESULT & DISCUSSION

The three-dimensional (3D) structure details of proteins are of major importance in providing insights into their molecular functions. Further analysis of 3D structures will help in the identification of binding sites and may lead to the designing of new drugs. The polyprotein of JE virus was obtained from the PIR protein database. Homology modeling is only a viable technique because it produces models that can be used for further research. Homology modeling helps in predicting the 3-D structure of a macromolecule with unknown structure (target) by comparing it with a known template (pdb|3I50|chainA) from another, structurally highly similar, macromolecule. The structure of the target protein is structurally similar with the template if both the target and template sequences are similar. In general, 30% sequence homology is required for generating useful models. Five 3D Models of our query protein was generated by the MODELER 9v7 (Figures 1, 2, 3, 4 and 5) and visualized by the Rasmol[29] with hydrogen bonds helices, strands and turns in the model (Table 3). Important physiochemical properties of the target protein were computed using Protparam tool to gain

an insight about the protein (Table 4). Polyprotein is a basic protein having 517 residues. Instability index of 25.75 indicates the stable nature of protein and a low Grand average of hydropathicity (GRAVY) value reflects its hydrated state.

Table 3
RasMol results of models.

S. No.	Features	Model-1	Model-2	Model-3	Model-4	Model-5
1.	Hydrogen Bonds	248	239	263	246	256
2.	Helices	7	6	5	5	6
3.	Strands	35	39	38	39	38
4.	Turns	63	64	58	64	66

Table 4
Important physiochemical properties of polyprotein determined using PROTPARAM.

S. No.	Property	Value
1.	Number of amino acids	517
2.	Molecular weight	55447.2
3.	Theoretical pl	7.67
4.	Total number of negatively charged residues (Asp + Glu)	44
5.	Total number of positively charged residues (Arg + Lys)	45
6.	Extinction coefficient	69620
7.	Extinction coefficient*	68870
8.	Instability index	25.75
9.	Aliphatic index	83.58
10.	Grand average of hdropathicity	0.029

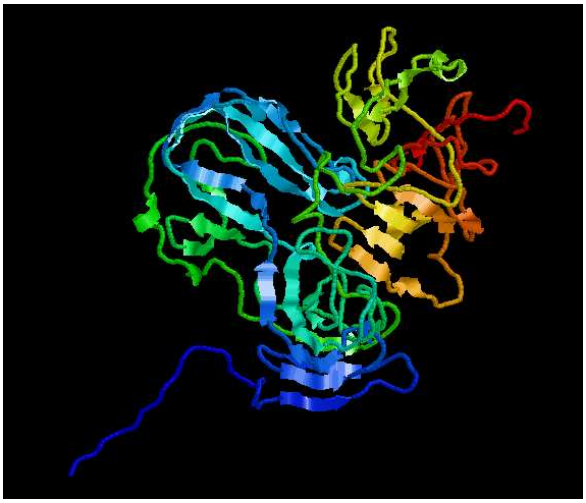

Figure 1
Ray diagram of 3D Model structure (C5NU03.B99990001)

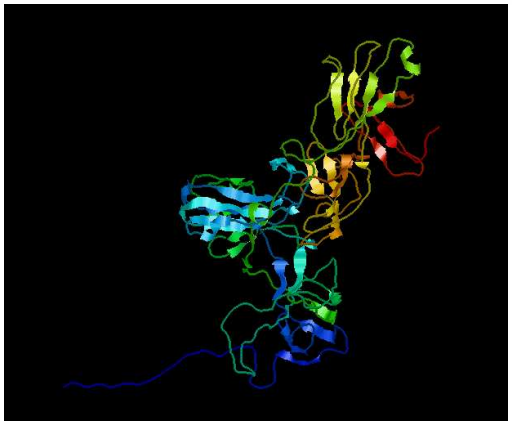

Figure 2
Ray diagram of 3D Model structure (C5NU03.B99990002)

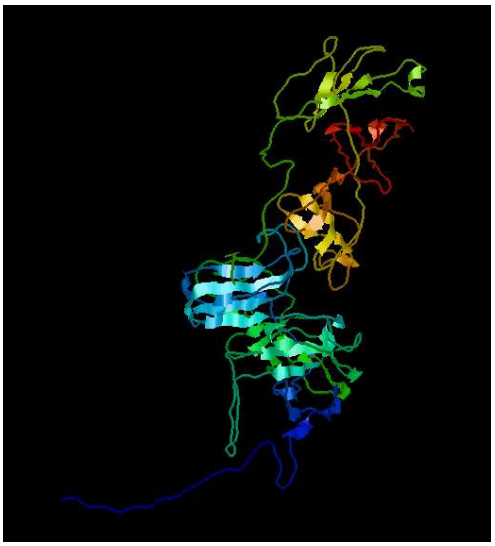

Figure 3
Ray diagram of 3D Model structure (C5NU03.B99990003)

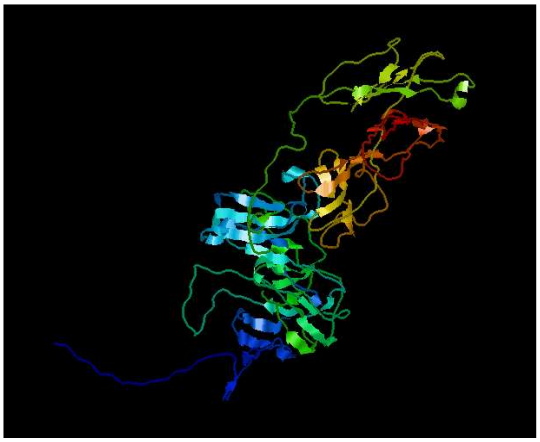

Figure 4
Ray diagram of 3D Model structure (C5NU03.B99990004)

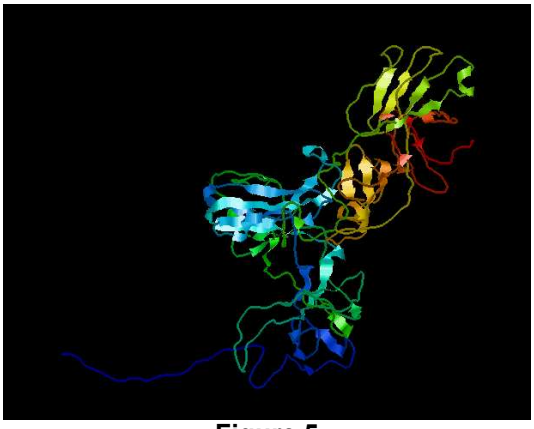

Figure 5
Ray diagram of 3D Model structure (C5NU03.B99990005)

The best model (Figure 1), is selected according to its highest molpdf & lowest Dope score (Table 2) and then verifying model with the help of WHATIF server and PROCHECK. After running the PROCHECK, Ramachandran plot shows (Figure 6) that 73.0% residues are in the favored region, 21.2% in the additional allowed region, and 2.7% in the generous allowed region and only 3.2% residues in the disallowed region which is acceptable and better than other generated models (Figures 2, 3 ,4 and 5).

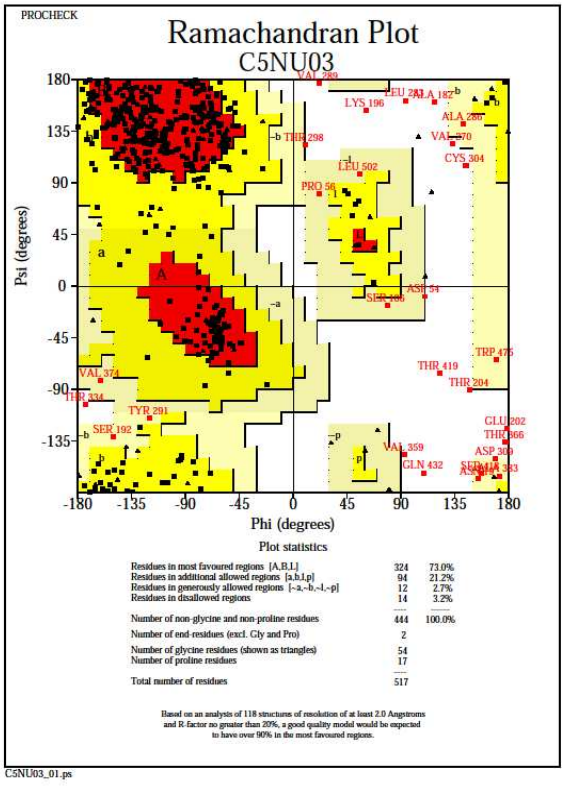

Figure 6
Ramachandran plot analysis of model using PROCHECK software

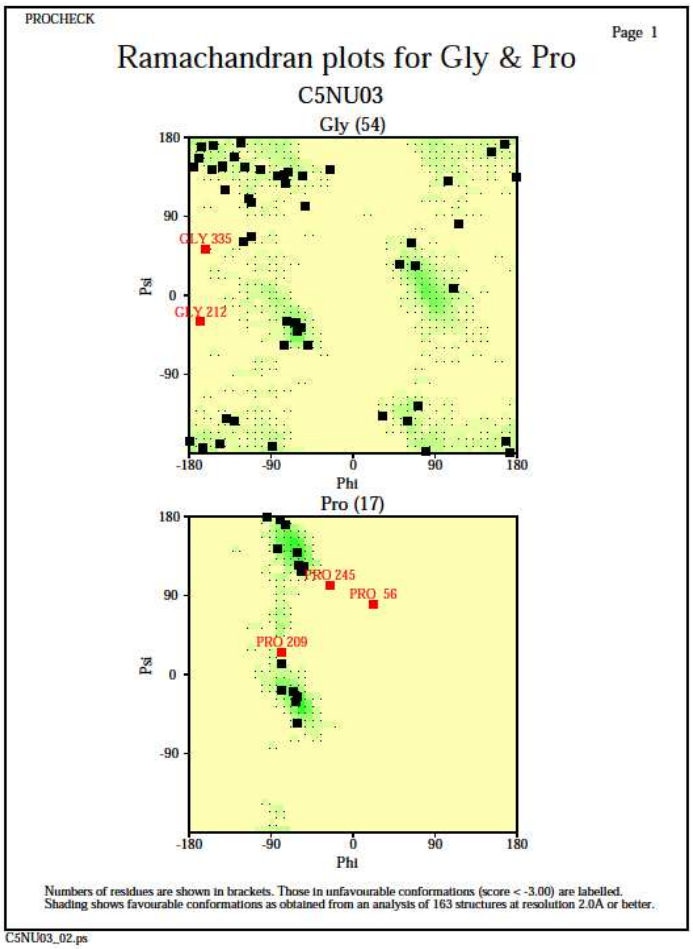

Figure 7
Ramachandran plot for Gly and Pro.

Ramachandran plot never shows the Gly and Pro, so these residues are shown in other plot (Figure 7). The plot shows that proline and glysine both are in allowed region. WHATIF server used to check the nomenclature of Torsion angles. The residues are sorted by residue type. From the tools VERIFY3D & ERRAT[30] which is available at Structural Analysis and Verification Server (SAVES) it was shown that 66.80% of the residues had an averaged 3D-1D score > 0.2 and overall quality factor 91%. The overall results provided the evidences that the predicted 3D structure of polyprotein is acceptable and of good quality. The model of the polyprotein will be very useful in wet laboratory while studying the real structure of the protein.

CONCLUSION

The result of comparative structural analysis shows that model (Figure 1) is the best structural model for polyprotein of Japanese Encephalitis Virus because it shows maximum residues (73.0%) in the favored region. *Insilico* study of protein helpful in almost all research fields, it not only saves money but also saves valuable time. Also we have predicted structure of polyprotein by homology modeling using different modeling software's. By the

computational methods we could predict the protein structure more rapidly and more economically. Now the predicted structure is ready to be verified by the in vitro analysis. Such model can be effectively used to structural information and can be further implemented in future drug designing.

ACKNOWLEDGEMENTS

This is not just to follow the custom of writing acknowledgement but to express and record my heartfelt feeling of gratefulness to all those who directly or indirectly helped me in this work. Further we wish to acknowledge protein homology modeling software's and other tools etc. for conducting this study

.

REFERENCE

1. Kuno, G, Chang, GJJ, Tsuchiya, KR, Karabatsos, N and Cropp, CB, Phylogeny of the genus Flavivirus. J Virol, 72, 73-83, (1998).
2. Monath TP, Levenbook I, Soike K, Zhang ZX, Ratterree M, Draper K, et al., Chimeric yellow fever virus 17D-Japanese encephalitis virus vaccine: dose-response effectiveness and extended safety testing in rhesus monkeys. J Virol, 74:1742-51, (2000). [PMID: 111650]
3. Solomon T, Nguyen MD, Kneen R, Gainsborough M, Vaughn DW and Khanh VT, Japanese encephalitis. J Neurol Neurosurg Psychiatry, 68: 405, (2000).
4. Monath TP and Heinz FX**,** Flaviviruses, p. 961–1034, ((1996). *In* B. N. Fields, D. M. Knipe, and P. M. Howley (ed.), Field's virology, 3rd ed. Lippincott-Raven Publishers, Philadelphia, Pa.
5. Sumiyoshi H, Mori C, Fuke I, Morita K, Kuhara S, Kondou J, Kikuchi Y, Nagamatu H and Igarashi A, Complete nucleotide sequence of the Japanese encephalitis virus genome RNA. Virology, 161: 497–510, (1987).
6. Vrati S, Giri RK, Razdan A and Malik P, Complete nucleotide sequence of an Indian strain of Japanese encephalitis virus: sequence comparison with other strains and phylogenetic analysis. Am J Trop Med Hyg, 61: 677–680, (1999).
7. Chambers TJ, Hahn CS, Galler R and Rice CM, Flavivirus genome organization, expression and replication. Annu Rev Microbiol, 44: 649–688, (1990).
8. Westaway EG, Mackenzie JM and Khromykh AA, Replication and gene function in Kunjin virus. Curr Top Microbiol Immuno,l 267: 323–351, (2002).
9. Chen HW, Pan CH, Liau MY, Jou R, Tsai CJ, Wu HJ, Lin YL and Tao MH, Screening of protective antigens of Japanese encephalitis virus by DNA immunization: a comparative study with conventional viral vaccines. J Virol, 73:10137-10145, (1999).
10. Kurane I, Immune responses to Japanese encephalitis virus. Curr Top Microbiol Immunol, 267 :91-103, (2002).
11. Mason PW, Dalrymple JM, Gentry MK, McCown JM, Hoke CH, Burke DS, Fournier MJ and Mason TL, Molecular characterization of a neutralizing domain of the Japanese encephalitis virus structural glycoprotein. J Gen Virol. 70(Pt 8):2037-2049, (1989).
12. Kaur R and Vrati S, Development of a recombinant vaccine against Japanese

encephalitis. J Neurovirol, 9:421-431, (2003).
13. Kojima A, Yasuda A, Asanuma H, Ishikawa T, Takamizawa A, Yasui K and Kurata T, Stable high-producer cell clone expressing virus-like particles of the Japanese encephalitis virus e protein for a second-generation subunit vaccine. J Virol, 77:8745-8755, (2003).
14. Konishi E, Fujii A and Mason PW, Generation and characterization of a mammalian cell line continuously expressing Japanese encephalitis virus subviral particles. J Virol, 75:2204-2212, (2001).
15. Wu HH, Chen CT, Lin YL and Lee ST, Sub-fragments of the envelope gene are highly protective against the Japanese encephalitis virus lethal infection in DNA priming – protein boosting immunization strategies. Vaccine, 22:793-800, (2004).
16. Konishi E, Pincus S, Paoletti E, Shope RE, Burrage T and Mason PW, Mice immunized with a subviral particle containing the Japanese encephalitis virus prM/M and E proteins are protected from lethal JEV infection. Virology, 188:714-720, (1992).
17. Chen LK, Lin YL, Liao CL, Lin CG, Huang YL, and Yeh CT, Persistence of Japanese encephalitis virus is associated with abnormal expression of the nonstructural protein NS1 in host cells. Virology, 223, 79-88, (1996).
18. Ting SHL, See E, Tan HC, Lee MA and Ooi EE, Development of a simplified assay for the detection of neutralizing antibodies to Japanese encephalitis virus. J Virol Methods, 93, 43-47, (2001).
19. Bernstein FC, Koetzle TF, Williams GJ, Meyer EFJ, Brice MD, Rodgers JR, Kennard O, Shimanouchi T and Tasumi M, The Protein Data Bank. A computer-based archival file for macromolecular structures. Eur J Biochem, 80(2): 319-24, (1977).
20. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN and Bourne PE, The Protein Data Bank. Nucleic Acids Res, 28 (1): 235-42, 2000.
21. Laskowski RA, MacArthur MW, Moss DS and Thornton JM, PROCHECK: A program to check the stereo chemical quality of protein structures. J Appl Cryst, 26: 283-291, (1993).
22. Marti-Renom MA, Stuart A, Fiser A, Sánchez R, Melo F and Sali A, Comparative protein structure modeling of genes and genomes. Annu Rev Biophys Biomol Struct, 29: 291-325, (2000).
23. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR and Appel RD et al. Protein identification and analysis tools on the ExPASy server. In The Proteomics Protocols Handbook (John M.,Walker), Humana Press, (2005).
24. Nybakken GE, Warren JT, Chen BR, Nelson CA and Fremont DH, Crystal structure of the West Nile Virus envelope glycoprotein in complex with the E53 antibody. J Embo Fab, 28: 3269-3276, (2009).
25. Fiser A and Sali A, Modeler: generation and refinement of homology-based protein structure models. Methods Enzymol, 374: 461-91, (2003).
26. Sali A and Blundell TL, Comparative protein modeling by satisfaction of spatial restraints. J Mol Bio, 234: 779-815, (1993).
27. Luthy R, James UB and Eisenberg D, Assessment of protein models with three-dimensional profiles. Nature, 356: 83-85, (1992).
28. Vriend G, WHATIF: A molecular modeling and drug design program. J Mol Graph, 8: 52–55, (1990).
29. Roger AS and Milner-White EJ, RasMol: Biomolecular graphics for all. Trends in Biochemical Sciences, 20: 374-376, (1995). RasMol was first widely distributed via the Internet in June, 1993, but this is the original paper publication describing RasMol.
30. Colovos C and Yeates TO, Verification of

protein structures: patterns of nonbonded atomic interactions. Protein Sci, 2(9):1511-1519, (1993).