Journal of Advanced Bioinformatics Applications and Research
ISSN 0976-2604.Online ISSN 2278–6007
Vol 5, Issue1, 2014, pp1-7
http://www.bipublication.com

HOMOLOGY MODELLING OF VP19 AND ITS VALIDATION FOR PREDICTION OF NOVEL BINDING SITES FOR WHITE SPOT SYDROME VIRUS

Sreejisha P.S[1] and Devika Pillai [2]

[1] M.E.S College for Advanced Studies, Edathala, Ernakulam, Kerala, India
[2] Department of Aquaculture, College of Fisheries,
Kerala University of Fisheries and Ocean Studies, Panangad, 682506, Kochi, Kerala, India
*Corresponding author: Email: devikamanoj.pillai@gmail.com, Tel: +91-9446111033.

[Received-12/10/2013, Accepted-11/01/2014]

ABSTRACT:

White spot syndrome (WSS) is one of most virulent viral disease known in the shrimp farming industry around the world. This is caused by a double stranded enveloped DNA virus-WSSV, the type species of the genus Whispovirus of the family Nimaviridae. White spot syndrome virus (WSSV) contains different proteins like envelope proteins, structural and nonstructural proteins. The envelope proteins are VP24, VP26, VP28 and VP19. VP19 is one of the foremost envelope protein causing the disease. These proteins are essential for entry into cells of the crustacean host. The aim of the present work is to generate the 3D structure of target protein VP19, validate the structure and predict its novel binding sites using bioinformatics techniques. The target protein was subjected to preliminary investigation by sequence analysis and similarity searching. The BLAST search of sequence against PDB (Protein Data Bank) database showed maximum homology with 3TB6_A. This was used as a template for structure prediction using SWISS PDB software. The newly generated structure was validated using Ramachandran plot. Bioinformatics tools are available for the detection of cavities in protein 3D structure and ranking them on the bases of probable binding clefts. The different validation steps proved that the protein structure is stable. Furthermore, the predicted structure can be selected for future studies thereby generating effective antiviral drugs.

Keywords: White Spot Syndrome Virus, VP19 protein, Homology modeling, Structure validation, Active site prediction.

[I] INTRODUCTION

White spot syndrome virus is a widely occurring highly virulent virus which attacks cultured shrimp and many other crustaceans [24], causing massive economic losses to the farmer. This virus induces distinctive clinical signs of white spots on the carapace, appendages, and the inner surface of the body [21]. The virus has an outer lipid bilayer membrane envelope, sometimes with

a tail like appendage at one end of the virion [25, 17]. The virion envelope contains two major proteins of VP28 and VP19 and the nucleocapsid consists of three major proteins of VP26, VP24 and VP15 [6, 19]. Study on the morphogenesis of the WSSV particle requires the genomic identification and chemical characterization of these WSSV virion proteins. An internal amino acid sequence of envelope protein VP19 was obtained by amino acid sequencing and used to locate the VP19 open reading frame of this protein on the genome, as WSSV ORF182. VP19 contained two putative transmembrane domains, which may anchor this protein in the WSSV envelope [23].

In this *in-silico* study, various computational tools were used for the primary examination by sequence analysis followed by homology modelling. Homology modelling of protein refers to constructing an atomic-resolution model of the "target" protein from its amino acid sequence and an experimental three-dimensional structure of a related homologous protein considered as "template". Homology modelling relies on the identification of one or more known protein structures likely to resemble the structure of the query sequence, and on the production of an alignment that maps residues in the query sequence to residues in the template sequence. The homology modelling procedure can be broken down into four sequential steps: template selection, target-template alignment, model construction, and model assessment.

[II] MATERIALS AND METHODS

2.1. Target Sequence Identification

In this *in silico* study, the selected target is an envelope protein VP19 based on its active involvement in causing white spot disease. The amino acid sequence of the target protein was retrieved from sequence database NCBI.

2.2. Primary Sequence Analysis

Various Bioinformatics tools were used to predict the primary sequence analysis. Study of physical and chemical properties, domain analysis, transmembrane helices, protein family prediction etc are the major areas in primary sequence analysis.

2.3. Secondary Structure Prediction

This technique aims to predict the secondary structures of protein based on knowledge of their primary structure. This means predicting the formation of protein structures such as alpha helices and beta strands.

2.4. Tertiary Structure Prediction and Validation

Homology modelling involves taking a known sequence with an unknown structure and mapping it against a known structure of one or several similar homologous proteins. It would be expected that two proteins of similar origin and function would have reasonable structural similarity. Therefore it is possible to use the known structure as a template for modelling the structure of the unknown structure. The stereo chemical validation of model structures of proteins is an important part of the comparative molecular modelling process.

2.5. Active Site Prediction

Active site is the small portion of a protein where substrate molecules bind and undergo a chemical reaction. Active site of the target protein was predicted using bioinformatics tools.

2.6. Structure Visualization

This program is aimed to display publication quality images of proteins. A molecular graphics visualization tool is required to view the structure that is encoded by atomic coordinate PDB files and to be able to manipulate the images to view the molecule from various perspectives.

[III] RESULTS AND DISCUSSION

The target sequence VP19 was obtained from the database NCBI with Accession no: ABG75925.1 which contains 121 amino acid residues.

3.1. Physical and Chemical Characterization

Various physical and chemical parameters of the target protein were calculated by PROTPARAM

server. The PROTPARAM tool computed parameters include 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) of the protein VP19 (Fig. 1).

3.2. Domain Prediction

Protein domain analysis is an essential component of protein sequence interpretation. Conserved domains are functional units within a protein that have been used as building blocks in molecular evolution. The bioinformatics tool, BLOCKS are multiply aligned ungapped segments corresponding to the most highly conserved regions of proteins. During the sequence entry in BLOCKS database, a set of blocks is converted to a position-specific score matrix. This matrix is used to score the motif found in the query sequence. Here the total number of BLOCKS was 152 (Fig. 2). The red color position shows the BLOCKS along the query sequence.

3.3. Transmembrane Helices Prediction

TMHMM program is used for the prediction of transmembrane helices in proteins. In addition to this, TMHMM discriminate between soluble and membrane proteins with high degree of accuracy. TMHMM are a membrane protein topology prediction method based on a hidden Markov model. The plot showed the posterior probabilities of inside/outside/TM helix (Fig. 3) of VP19. It predicts two transmembrane helices of VP19, which probably helps the protein to hang on the viral envelope.

3.4. Protein Family Prediction

This helps to predict the protein families which are similar to the query sequence. Similar families should have identical functions. Pfam is used for the prediction of similar families. The Pfam database is a large collection of protein families, each represented by multiple sequence alignments and hidden Markov models (HMMs). HMM scan calculates the matching score between the query sequence and each domain found in Pfam library in bit score and in E-value. Pfam result showed that the protein VP19 has nine similar families (Fig. 4).

3.5. Secondary Structure Analysis

SOPMA (Self Optimized Prediction Method with Alignment) is a secondary structure prediction program that uses multiple alignments. SOPMA has been described to improve the success rate in the prediction of the secondary structure of proteins. This predicts 69.5% of amino acids for a three-state description of the secondary structure such as alpha-helix, beta-sheet and coil. The result of the present study showed 31.40% alpha helix, 25.62% extended strand, 41.32% random coil and 1.65% beta turns in target protein (Fig. 5).

3.6. Homology Modelling: Similarity Searching

Similarity searching is the initial step of homology modelling. The simplest method of template identification relies on serial pairwise sequence alignments by BLAST. In BLAST the target sequence searched similarity against the structure database PDB. BLAST search helped to select the best hit based on maximum identity with low E-value. The selected template structure showed 37% of similarity with query sequence (Fig. 6).

3.7. Template Selection

Choosing the best template among the candidates from BLAST result is a key step. The selected template 3D structure was downloaded from template library of SWISS-MODEL server. Its PDB ID is 3TB6_A (Fig. 7).

3.8. Target-Template Alignment

To build a model, all comparative modelling programs depend on a list of assumed structural equivalences between the target and template residues. This list is defined by the alignment of the target and template structure. Here the query sequence of VP19 was aligned with the template

3TB6_A. This homology modeling technique was carried out by SWISS-PDB software (Fig. 8).

3.9. Model Refinement

This is the final stage of homology modeling. Three-dimensional structural model of the target was generated (Fig. 9).

3.10. Structure Validation

Structure validation is the process of evaluating reliability for 3-dimensional atomic models of target protein VP19. These models provide 3D coordinates for each atom in the protein. PROCHECK helped to check the stereo chemical quality of a protein structure by analyzing residue-by-residue geometry and overall structure geometry (Fig. 10). As a result of the Ramachandran Plot analysis, we proved that the target protein is suitable for further studies. The Ramachandran Plot showed the residues to be present in the most favored regions as 91.6%.

3.11. Energy Minimization

Energy minimization is an important step in molecular modeling with applications in molecular docking and in mapping binding sites. Minimization involves repeated evaluation of various bonded and non-bonded energies of a protein complex. Energy minimization can repair distorted geometries by moving atoms to release internal constraints. Atoms should be displaced in order to reach a lower energy state are shown by dotted lines (Fig. 11).

3.12. Structure Visualization

The molecular visualization tool Rasmol was used to display the modeled protein molecule on the screen in a variety of color schemes and molecule representations (Fig. 12). The program reads in a molecule co-ordinate file and interactively displays the protein structures in different ways such as wireframe, backbone, sticks, spacefill, ribbon etc.

3.13. Active site Prediction

Active site is found in a pocket that is lined by amino acid residues that participate in recognition of the substrate. POCASA is an automatic program which can predict binding sites by detecting pockets of proteins of known 3D structure. The target protein VP19 showed 2 pockets in different colors (Fig. 13).

Number of amino acids: 121

Molecular weight: 13241.9

Theoretical pI: 4.36

Amino acid composition: CSV format

Ala (A)	12	9.9%
Arg (R)	6	5.0%
Asn (N)	5	4.1%
Asp (D)	16	13.2%
Cys (C)	0	0.0%
Gln (Q)	2	1.7%
Glu (E)	3	2.5%
Gly (G)	8	6.6%
His (H)	0	0.0%
Ile (I)	4	3.3%
Leu (L)	11	9.1%
Lys (K)	6	5.0%
Met (M)	10	8.3%

Figure 1: ProtParam result

Blocks Name:
IPB003298O
Description:
Apical membrane antigen 1 signature
Appearance (Score):

Position	Sequence	Score
76..100	DdtAndnddEdkyknrTrdmMllag	1109
77..101	DtandnddEDKyknrTrdmmllags	1081
26..50	DlEgSmsmarmglflivaisigilv	1067
67..91	DtdkdtdddDdtAndnddedkyKnr	1031
94..118	DmmllagsallflvsaatvfMsyPk	1028
17..41	aagpSytmEDlegSmsmarmglfli	1027
83..107	DdEdkyknrtrdmmllagsallflv	1020

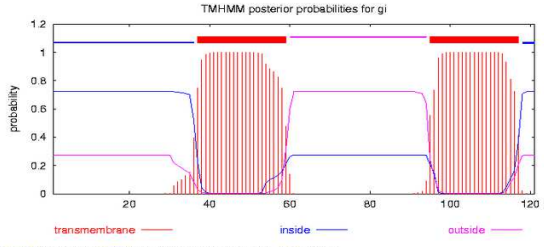
Sequence:
MATTTNTLPFGRTGAQAAGPSYTMEDLEGSMSMARMGLFLIVAISIGILVLAVMNVWMGP
KKDSDSDTDKDTDDDDDTANDNDDEDKYKNRTRDMMLLAGSALLFLVSAATVFMSYPKRR
Q

Figure 2: Blocks result

TMHMM result

HELP with output formats

```
# gi Length: 121
# gi Number of predicted TMHs:  2
# gi Exp number of AAs in TMHs: 43.68573
# gi Exp number, first 60 AAs:  22.43477
# gi Total prob of N-in:        0.72431
# gi POSSIBLE N-term signal sequence
gi      TMHMM2.0        inside     1    36
gi      TMHMM2.0        TMhelix    37   59
gi      TMHMM2.0        outside    60   94
gi      TMHMM2.0        TMhelix    95   117
gi      TMHMM2.0        inside     118  121
```



plot in postscript, script for making the plot in gnuplot, data for plot

Figure 3: TMHMM result

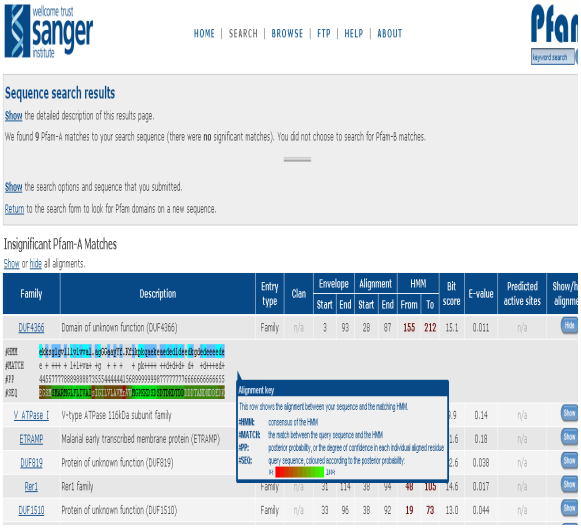


Figure 4: Pfam result

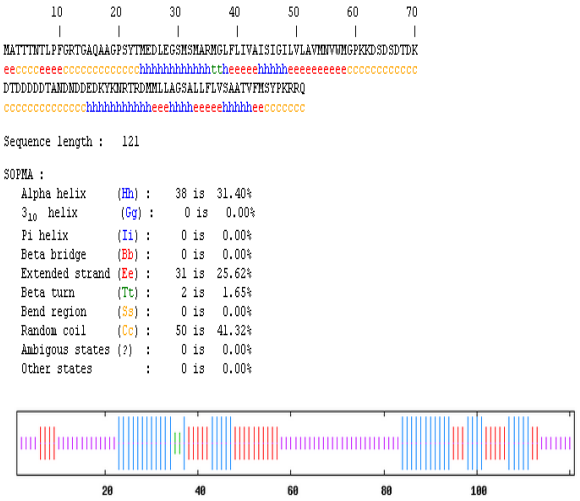


Figure 5: SOPMA result

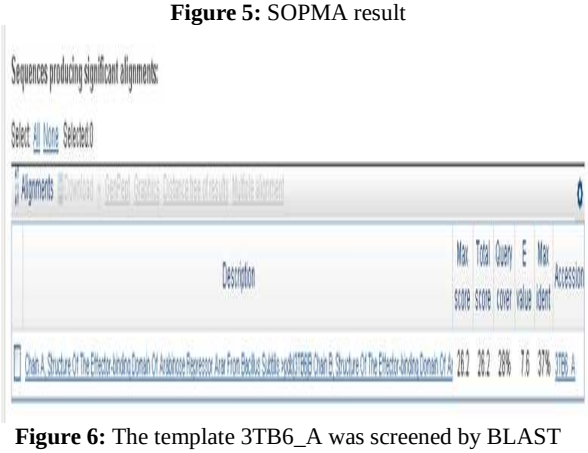


Figure 6: The template 3TB6_A was screened by BLAST

Figure 7: Selected template structure

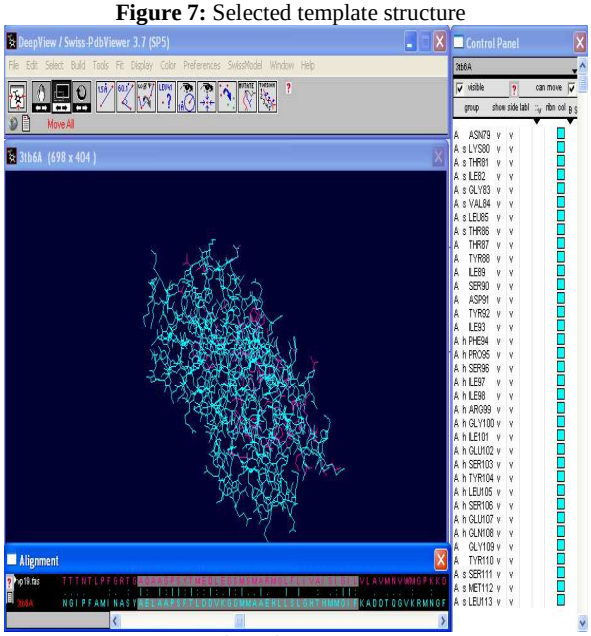


Figure 8: Target-Template alignment in Swiss-PDB software

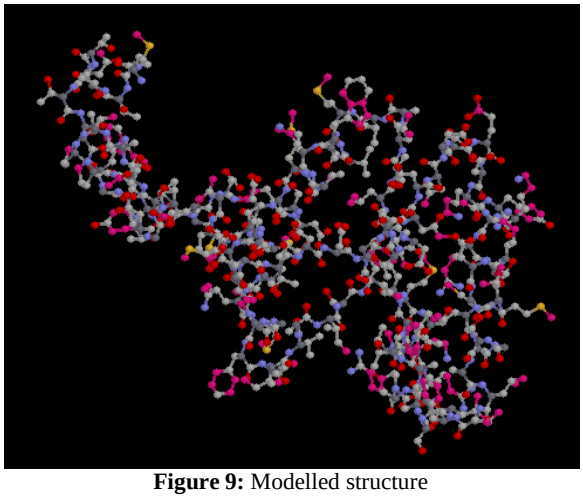


Figure 9: Modelled structure

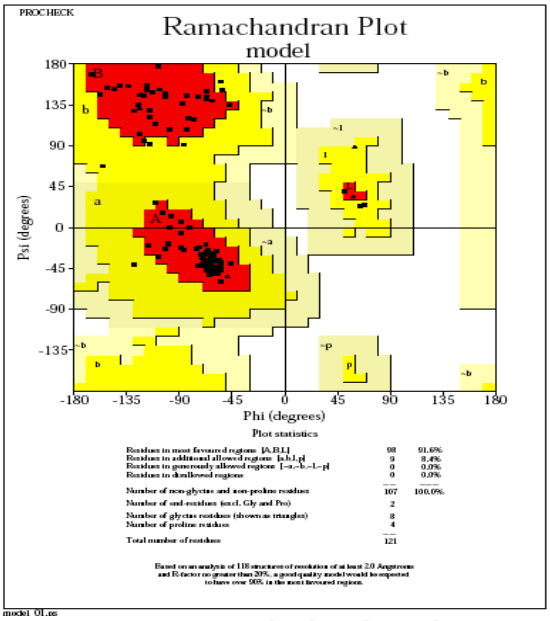


Figure 10: Ramachandran Plot Analysis

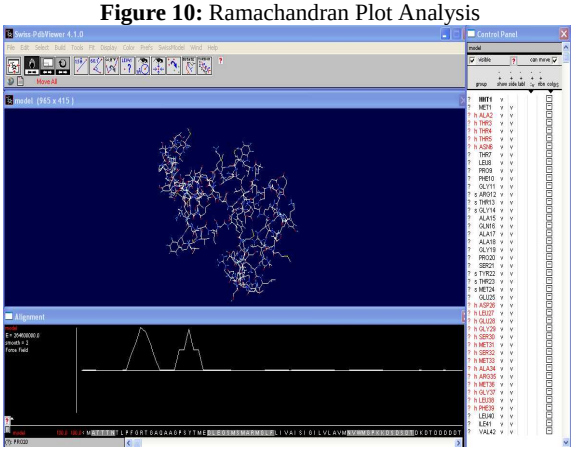


Figure 11: Energy minimized in Swiss-PDB software

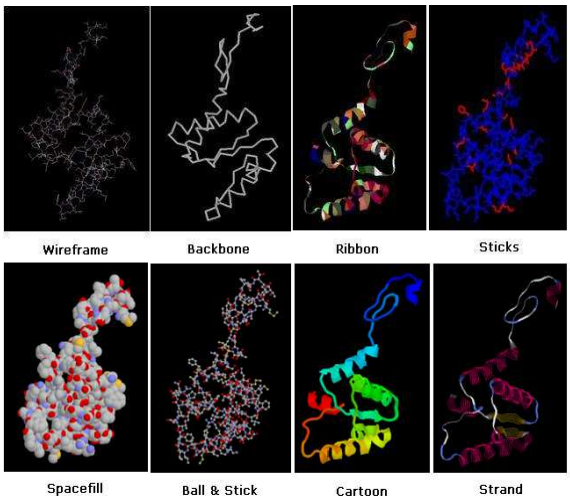


Figure 12: Rasmol result

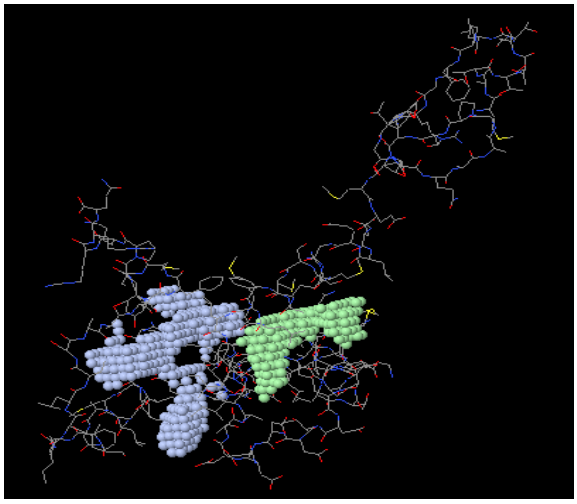


Figure 13: Two pockets are found in VP19 structure

[V] CONCLUSION

White spot syndrome virus (WSSV) is one of the most serious pathogens of the cultured shrimp. In WSSV the protein VP19 plays a crucial role for the spreading of the disease. Based on this literature knowledge, we have selected VP19 as a target protein in this study. But the 3D structure of VP19 protein was not available in any structure database. So we initiated this *in-silico* method to generate the structure of VP19. The target protein showed a successive result in initial studies by sequence analysis and similarity searching. BLAST search sequence indicated maximum homology with template and structure prediction was carried out by SWISS-PDB software. Newly generated structure was validated and also its active sites predicted. These studies suggest that the protein structure is stable. Furthermore, the predicted structure can be selected for future studies thereby generating effective antiviral drugs.

ACKNOWLEDGEMENT

This work was supported by Kerala University of Fisheries and Ocean Studies (KUFOS), Panangad, Kochi, Kerala, India.

REFERENCES

1. Altschul, S.F, et.al., (1990), Basic local alignment search tool, J. Mol. Biol. 215, 3, 403-10
2. Arnold, K, et.al., (2006), The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling, Bioinformatics. 22, 195-201
3. Awale, M. K, et.al., (2010), Homology modeling and atomic level binding study of Leishmania MAPK with inhibitors, J. Mol. Model. 16, 475–488
4. Bendtsen, J.D, et.al., (2004), Improved prediction of signal peptides: SignalP 3.0. J Mol Biol. 340, 783–795.
5. Berman, H.M. et.al., (2000), The protein data bank, Nucleic Acids Res. 28, 235–242
6. Chen, L, et.al., (2002), Identification of a nucleocapsid protein (VP35) gene of shrimp white spot syndrome virus and characterization of the motif important for targeting VP35 to the nuclei of transfected insect cells, J. Virol. 29, 344-353
7. Colovos, C, et.al., (1993), Verification of protein structures: patterns of nonbonded atomic interactions, Protein Sci. 2, 1511–1519
8. Eswar, N, et.al., (2003), Tools for comparative protein structure modeling and analysis, Nucleic Acids Res. 31, 3375–3380
9. Gasteiger, E, (2005), Protein Identification and Analysis Tools on the ExPASy Server. In: John M. Walker, The Proteomics Protocols Handbook, Humana Press. 571-607.
10. Geourjon, C, et.al., (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments, Comput Appl Biosci. 11: 681-684.
11. Gupta, S, et.al., (2011), Prediction of a new surface binding pocket and evaluation of inhibitors against huntingtin interacting protein 14: an insight using docking studies, J. Mol. Model. 17, 3047–3056
12. Hendrik, M, et.al., (2006), *In-silico* identification of putative promoter motifs of White Spot Syndrome Virus, BMC Bioinformatics. 7, 309
13. Hirokawa, T, et.al., (1998), SOSUI: classification and secondary structure prediction system for membrane proteins, Bioinformatics. 14, 378–379
14. Huang, R, et.al., (2005), A novel envelope protein involved in White spot syndrome virus infection, J. Gen. Virol. 86, 1357-1361
15. Krogh, A, et.al., (2001), Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, J MolBiol. 305, 567– 580.
16. Laskowski, R.A, et.al., (1993), PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Cryst. 26, 283–291
17. Nadala, E, et.al., (1998), Characterization of a non-occluded baculovirus-like agent pathogenic to penaeid shrimp, Dis. Aquat. Org. 33, 221-229
18. Ode, H, et.al., (2012), Molecular dynamics simulation in virus research, Front Microbiol. 3, 258
19. Qiang, W**,** et.al., (2008), VP26 of White Spot Syndrome Virus Functions as a Linker Protein between the Envelope and Nucleocapsid of Virions by Binding with VP51, J. Virol. 82, 12598–12601
20. Rajeev, K, et.al., (2005), Production of recombinant enveloped structural proteins from the Chinese WSSV isolate, Indian J. Clin. Biochem. 20, 136–141
21. Stephanie, D, et.al., (1997), Ultrastructure and morphogenesis of White Spot Syndrome Baculovirus (WSSV), Dis Aquat Org. 29, 205-211
22. Subramaniam, S, et.al., (2009), Molecular modeling studies of the interaction between Plasmodium falciparum HslU and HslV subunits, J. Biomol. Struct.Dyn. 26, 473–479
23. VanHulten, M.C, et.al., (2002), Identification of VP19 and VP15 of white spot syndrome virus (WSSV) and glycosylation status of the WSSV major structural proteins, J. Gen. Virol. 83, 257-65
24. Wenlin, Wu, et.al., (2002), Identification of white spot syndrome virus (WSSV) envelope proteins involved in shrimp infection, J. Virol. 332, 578 – 583
25. Witteveldt, J**,** et.al., (2004), Protection of *Penaeusmonodon* against White Spot Syndrome Virus by Oral Vaccination, J. Virol. 78, 2057–2061