In Silico Design and Homology Modeling of Helicase C-Terminal Domain of Nonstructural Protein NS3 of West Nile Virus (Strain NY-99)

Shamrat Kumar Paul, Abul Bashar Ripon Khalipha, Lutful Kabir, Pranta Ray, Masuma Akter Mina and Hossain Mohammad Masum

Abstract West Nile virus (WNV) is a mosquito-transmitted single-stranded (ss) RNA flavivirus which causes West Nile infection and human disease of variable severity. Most of the people infected with WNV either have only minor signs or symptoms such as mild headache and fever or do not develop any signs or symptoms. Sometimes WNV causes a life-threatening illness as well as inflammation of the spinal cord or brain in infected patients. The genome of WNV is single-stranded RNA (ssRNA) that contains the characteristics of positive polarity (mRNA sense). The strains of WNV are classified into two groups, such as lineage (1 and 2). The basis of grouping these strains based on the substitutions and deletions of signature amino acid in the envelope protein sequence. Among them, Lineage 1 WNV strains are responsible for developing the West Nile infection in humans. The genome of WNV consists of a single open reading frame (ORF) and it produced ten mature viral proteins which are classified into structural and nonstructural. Nonstructural protein 2B (NS2B) and Serine Protease (NS3) are two of nonstructural proteins of them. NS2B-NS3 has a major role in proteolytic processing of these nonstructural proteins. NS3 has several enzymatic activities like NTPase and RNA helicase. Among them, the helicase activities is one of them. There is evidence that Helicase C-terminal domain of West Nile Virus has a major role in the unwinding of dsRNA. But Helicase C-terminal domain is not yet explored, so our purpose is to investigate the physicochemical, structural, and functional features of Helicase C-terminal domain. Molecular modeling of the unexplored Helicase C-terminal domain was generated by using Phyre2 and Swiss Model. The prediction of active ligand binding sites is generated by using PredictProtein server. This results that Helicase C-terminal

S. K. Paul · L. Kabir · M. A. Mina
Department of Biochemistry and Molecular Biology, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, Bangladesh

A. B. R. Khalipha (✉) · P. Ray
Department of Pharmacy, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, Bangladesh
e-mail: khalipha1982@gmail.com

H. M. Masum
Institut National de La Recherche Agronomique, Paris, France

domain protein is stable and its nature is acidic which thermostability is high and has better hydrophilic property. By the putative transferase and catalytic functional activity, 14 binding sites were predicted. In Homology Modeling, it is revealed that 14 binding sites are predicted as ligand binding sites. This investigation decoded the role of this unexplored Helicase C-terminal domain of West Nile virus (WNV), and so it can pave the way for enriching our knowledge for pathogenesis and medication of WNV infection.

1 Introduction

In the middle of 1999, this mosquito-borne zoonosis was first identified in the United States. More than 39,000 reported clinical cases, of which 17,463 were presented with neuroinvasive disease, 22,094 with West Nile fever, and 1668 with a dangerous condition of patients caused by it [1]. Mosquitoes obtain this virus by biting the infected birds and this virus is transmitted to a human when that infected mosquitoes bite on human [2]. Characteristically the WNV genome is a single-stranded RNA of positive polarity mRNA sense. At the 5′ end there is a type 1 cap structure but at the 3′ end, it terminates with CUOH because compared with other viruses Flavivirus genomes are the only mammalian plus-strand RNA virus genomes which do not contain any 3′ Poly A tail [3]. The length of the WNV genome is 11,029 nt and it has only a single open reading frame (ORF) of 10,301 nt. In its single ORF, there is structural and nonstructural part [4]. From these parts first, it produces a single polyprotein and by proteolytic processing of viral serine protease (NS2B-NS3) and various cellular proteases that polyprotein turns into ten mature viral protein [5]. From the structural part which is situated at the 5′ portion it produces three viral structural proteins, capsid (C), membrane (prM/M), and envelope (E) and the rest seven nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) are encoded from the 3′ nonstructural portion [6]. Recently, the molecular study by Brinton, M. A. [7] suggested that the NS2B and NS3 combinedly act as serine protease which has a great role in proteolytic processing of that single polyprotein. Furthermore, it uptakes energy from NTP hydrolysis for RNA unwinding by Helicase activity. The literature review reveals that in the NS3 protein, it has Helicase C-terminal Domain which has a crucial role in this process [8]. However, the structure of this Helicase C-terminal Domain is not reported yet. Also, the detailed physicochemical characterization and putative structure with ligand binding active sites are not elucidated, so we proposed an in silico 3-D structure prediction of Helicase C-terminal Domain using homology modeling.

2   Methodology

2.1   Retrieval of the Target Amino Acid Sequence of Helicase C-Terminal Domain

The amino acid sequence of Helicase C-terminal domain of NS3 (Genome Polyprotein-GP1) was obtained from UniProtKB with the ID Q9Q6P4. Due to the unavailability of 3-D structure in PDB, modeling of this unexplored domain was undertaken utilizing 166 aa long sequence of Helicase C-terminal domain of West Nile virus.

2.2   Physicochemical Characterization

Physicochemical properties of the retrieved sequence were determined using two web-based servers. ProtParam tool of Expasy (http://us.expasy.org/tools/protparam. html) employed for the prediction of amino acid composition, instability and aliphatic indices, extinction coefficients, and grand average of hydropathicity (GRAVY) [9]. Theoretical isoelectrical point (pI) was calculated using Sequence Manipulation Suite (SMS) Version 2 (http://www.bioinformatics.org/sms2/protein_iep.html).

2.3   Secondary Structure Prediction of Helicase C-Terminal Domain Protein

The self-optimized prediction method with alignment (SOPMA; https://npsa-prabi. ibcp.fr/NPSA/npsa_sopma.htm) software [10] and (PSIPRED program; http://bioinf. cs.ucl.ac.uk/psipred) was used to predict the secondary structure of Helicase C-terminal Domain protein (target). Disorder prediction was performed using DISOPRED tool (Predict Protein server; https://predictprotein.org) used to predict secondary structure [11].

2.4   Homology Modeling and Validation of Helicase C-Terminal Domain

There is no experimentally deduced 3D structure available for Helicase C-terminal domain protein in protein data bank (PDB); therefore, homology modeling of the protein of Helicase C-terminal domain was done using two programs (Swiss Model and Phyre2; http://www.sbg.bio.ic.ac.uk/phyre2) [12, 13]. Secondary structure has

also been predicted using Phyre2. 3D model of Helicase C-terminal domain generated from Swiss-Model and Phyre2 was compared and only the most suitable 3D model was selected for final validation. The final modeled structure was validated using Ramachandran plot analysis (PROCHECK; http://nihserver.mbi.ucla.edu/SAVES) for stereochemical property.

3   Result and Discussion

The amino acid sequence of Helicase C-terminal domain was retrieved in FASTA format and used as query sequence for determination of physicochemical parameters. The instability index of 37.18 (<40) indicated the stable nature of Helicase C-terminal domain [30]. The protein is acidic in nature (pI 6.23, 5.82*) with molecular weight of the 18.66 kDa. High extinction coefficient values (28,670) indicate the presence of Cys, Trp, and Tyr residues [14]. Higher aliphatic index values (70.57) of the query protein suggested as a positive factor for increased thermos-stability for a wide temperature range [15]. Hydrophilic nature of the protein and the possibility of better interaction with water [16] were indicated by the lower grand average of hydropathicity GRAVY indices value (−0.319) as shown in Table 1. The default parameters (similarity threshold: 8; window width: 17) were considered by SOPMA for the secondary structure prediction with >70% prediction accuracy. Utilizing 511 proteins (sub-database) and 15 aligned proteins, SOPMA predicted 40.96% of residues as random coils in comparison to Alpha helix (25.90%), extended strand (24.10%) and Beta turn (9.04%) as shown in Table 2. PSIPRED showing the higher confidence of prediction of helix, strand and coil (Fig. 1). Secondary structure prediction by PredictProtein employing neural network system, provides the prediction accuracy of more than 72%. 22.9% helix conformation (α; π; 3_10-helix), 44.6% loop (L)

Table 1 Physicochemical parameters computed using Expasy's ProtParam and SMS tool

Physicochemical parameters	Values
No. of amino acid (aa)	166
Molecular weight (MW)	18658.94
Theoretical isoelectric point (pI)	6.23, 5.82*
Aliphatic index	70.57
Instability index	37.18
Extinction coefficient (All Cys form Cysteine)	28,670
Extinction coefficient (All Cys reduced)	28,420
Total no. of negatively charged residues (Asp + Glu)	17
Total no. of positively charged residues (Arg + Lys)	15
GRAVY (Grand average of hydropathicity)	−0.319

*pI determined by SMS Version2

Table 2 Secondary structure elements prediction by SOPMA

Secondary structure elements	Values (%)
Alpha helix (Hh)	25.90
310 helix (Gg)	0.00
Pi helix (Ii)	0.00
Beta bridge (Bb)	0.00
Extended strand (Ee)	24.10
Beta turn (Tt)	9.04
Bend region (Ss)	0.00
Random coil (Cc)	40.96
Ambiguous states (?)	0.00

followed by 32.5% beta strand (E = extended strand in beta sheet conformation) was predicted in Helicase C-terminal domain. Intrinsic disorder profile was computed using DISOPRED and >90% of the amino acid is below the confidence score of 0.5 for disordered condition, suggested the lowest possibility of distortion and conferred the high stability to the predicted protein.

Binding sites were predicted using predict protein server, where 14 different protein binding sites were identified at positions, viz.: 30–31; 41; 43; 52–54; 81; 83; 92–94; 104; 110–114; 114; 117; 120–122; 124–124; and 133–136 (Fig. 2). Gene ontology predicted and categories the functional aspects as cellular, molecular, and biological, where this Helicase C-terminal domain protein found to be subcellular or the part of host cell or membrane; metabolic processes such as primary and cellular metabolic processes including cyclic, heterocyclic, and aromatic compound metabolism processes. Molecular function including binding (Score: 49) involves heterocyclic (Score: 49), organic (49), cyclic compound binding; small molecule (Score: 49), and nucleotide binding (Score: 49) whereas and nucleic acid binding (Score: 32) include nucleoside-triphosphatase activity (Score: 32) showed in Table 3.

Helicase C-terminal domain target sequence was inserted as input (FASTA format) in Swiss-Model workspace. The Swiss-MODEL template library (SMTL) was searched with HHBlits [17] resulted in total 32 templates. Among the seven most favorable template (2qeq.1.A; 2qeq.1.A; 2qeq.2.A; 2qeq.2.A 2z83.1.A; 2z83.1.A; 2v80.1.A) are those who have GMQE value greater than 0.90. Target sequence was selected based on the Qualitative Model Energy Analysis (QMEAN) score (−1.33), Global model quality estimate (GMQE) 0.96, percentage of sequence identity (96.39), similarity (0.61) and coverage (1.00). The model was based on target template alignment using ProMod3, where insertion, deletions remodeled, and side chains were then rebuilt.

Our model showed resemblance with 2qeq.1.A. So that ProMod3 Program selected 2qeq.1.A. as template structure which belongs to Flavivirin protease NS3 catalytic subunit. Template list for homology modeled structure is showed in Table 4. The model was generated and saved in PDB format (Fig. 3 left). PROCHECK, another

Fig. 1 Secondary structure prediction by PSIPRED

Fig. 2 Protein–protein and protein-polynucleotide binding sites

tool was used to measure the accuracy of protein models. Ramachandran plot statistics of Helicase C-Terminal revealed that 91.7% of amino acid residues were found in the most favored regions. Thus, the selected protein model was very good as all of the amino acid residues were within the limits of Ramachandran plot (Fig. 5). Verification was also done by ERRAT and QMEAN server. The respective values Z-scores of

Table 3 Molecular functional ontology

GO id	GO term	Reliability (%)
GO:0005488	Binding	49
GO:1901363	Heterocyclic compound binding	49
GO:0000166	Nucleotide binding	49
GO:1901265	Nucleoside phosphate binding	49
GO:0036094	Small molecule binding	49
GO:0097159	Organic cyclic compound binding	49
GO:0003676	Nucleic acid binding	36
GO:0003723	RNA binding	36
GO:0016787	Hydrolase activity	32
GO:0017111	Nucleoside-triphosphatase activity	32

Table 4 Alignment of selected template

Name	Title	GMQE	Identity (%)
2qeq.1.A	Flavivirin protease NS3 catalytic subunit	0.98	96.39
2qeq.2.A	Flavivirin protease NS3 catalytic subunit	0.77	96.39
2qeq.1.A	Flavivirin protease NS3 catalytic subunit	0.97	96.39
2qeq.2.A	Flavivirin protease NS3 catalytic subunit	0.63	96.39
2z83.1.A	Helicase/Nucleoside-triphosphatase	0.96	84.94
2z83.1.A	Helicase/Nucleoside-triphosphatase	0.94	84.85
2v8o.1.A	Flavivirin protease NS3	0.95	80.12
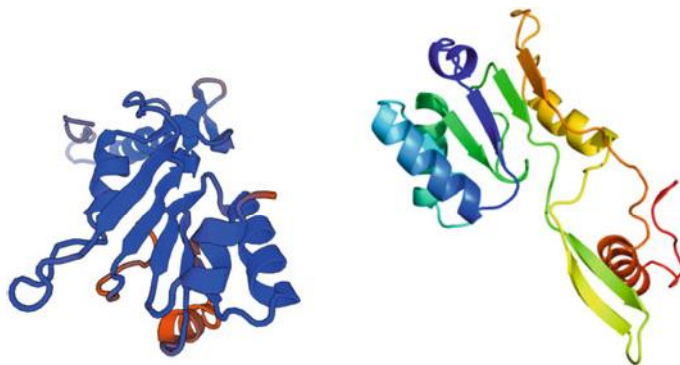
Fig. 3 Helicase C domain structure with helix, strands and coil predicted by Swiss Model [left], Helicase-C domain structure predicted by Phyre2 [right]

Cbeta interaction energy, torsion angle energy, solvation energy, secondary structure in case of Helicase C-Terminal Domain are −1.89, −0.79, −0.37, and −0.56. The overall QMEAN score for Helicase C-Terminal is −1.43. QMEAN generated results confers Helicase C-Terminal Domain as a qualified model for drug target scopes. Stereochemical quality of the Swiss model predicted Helicase C-terminal domain structure was evaluated by plotting Ramachandran map (PROCHECK) shown in Fig 5. 91.7% of the total residues (132) (Table 5) were found in the core (A; B; L; P) whereas 7.6% of residues were in the allowed (a; b; l; p) regions. Disallowed region constitutes 0.0% of the residues. PROCHECK analysis showed max deviation of 4.2 (residue properties), with bond length/angle of 5.1 and 93.9% planar groups within the limits (Fig. 4).

Similarly, the homology modeling of Helicase C-terminal domain was performed by Phyre2. Based on the six templates (d1yksa2, c2vbcA, c2wv9A, c2jlrA, c5aorA, and c5vheA), protein model was generated with 95% of the residues modeled at 100% confidence with the template structure d1yksa2 which belongs to Crystal structure of yellow fever virus NS3 helicase and the homology modeled protein is shown in Fig. 3 right with coordinates (A): X: 56.637, Y: 44.638, Z: 56.329 (based

Table 5 Ramachandran plot analysis

Ramachandran plot statistics	Residue	%
Residues in the most favored regions [A, B, L]	132	91.7
Residues in the additional allowed regions [a, b, l, p]	11	7.6
Residues in the generously allowed regions [a, b, l, p]	1	0.7
Residues in the disallowed regions [xx]	0	0.00
Number of non-glycine and non-proline residues	144	100.0
Number of end residues (excl. Gly and Pro)	1	–
Number of glycine residues	14	–
Number of proline residues	7	–
Total number of residues	166	–

Fig. 4 Homology modeled hypothetical Helicase C-terminal domain visualized by discovery studio
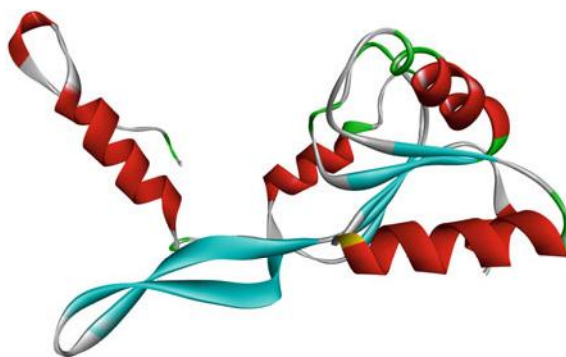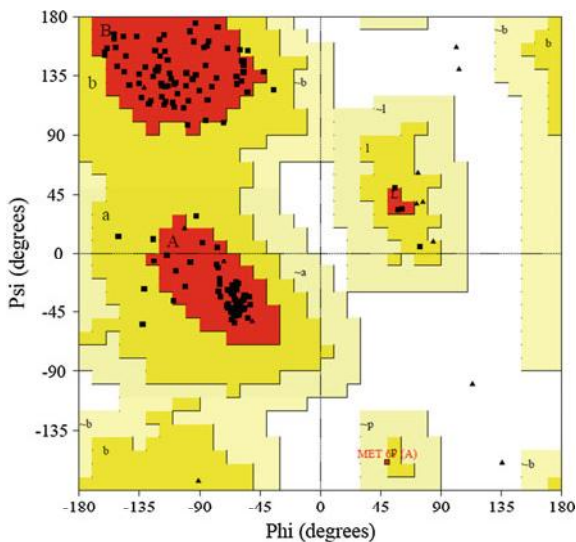
Fig. 5 Ramachandran plot analysis of Helicase C-terminal through Swiss Model workshop

on heuristics to maximize confidence, percentage identity, and alignment coverage). Secondary structure prediction by Phyre2 was described as Disordered (22%), Alpha helix (32%), and beta strand (25%) (data not shown). Phyre2 predicted structural model was evaluated for the stereochemical quality using Ramachandran map (PROCHECK). 88.8% of the residues were found in the core (A; B; L; P), whereas 8.4% of residues were in the allowed (a; b; l) regions. However, 2.8% residues were aligned in generously allowed region (~a, ~b, ~l ~p), whereas disallowed region constituted 1.5% of the residues. Among residual properties max deviation was 0.00%, bond length/angle 18.2 with 1 cis-peptides with 92.4% planar groups within limits (data not shown).

4 Conclusion

We have reported the structural model of Helicase C-terminal domain with predicted active site for ligand binding through in silico approach whereas we showed the position of amino acid in favored region by evaluating the structure. The physico-chemical parameters prediction and functional annotation are useful for understanding the action of this viral protein domain activity. Our homology-modeled protein domain provides insights into the functional role of Helicase C-terminal domain in viral pathogenesis which will help to design potential therapeutic drug against West Nile Infection.

References

1. Lanteri, M.C., Lee, T.H., Wen, L., Kaidarova, Z., Bravo, M.D., Kiely, N.E., Kamel, H.T., Tobler, L.H., Norris, P.J., Busch, M.P.: West Nile Virus nucleic acid persistence in whole blood months after clearance in plasma: implication for transfusion and transplantation safety. Transfusion **54**(12), 3232–3241 (2014)
2. Hamer, G.L., Kitron, U.D., Brawn, J.D., Loss, S.R., Ruiz, M.O., Goldberg, T.L., Walker, E.D.: Culex pipiens (Diptera: Culicidae): a bridge vector of West Nile virus to humans. J. Med. Entomol. **45**(1), 125–128 (2008)
3. Liu, Y., Wimmer, E., Paul, A.V.: Cis-acting RNA elements in human and animal plus-strand RNA viruses. Biochimica Et Biophysica Acta (BBA)-Gene Regul. Mech. **1789**(9–10), 495–517 (2009)
4. Yu, L., Nomaguchi, M., Padmanabhan, R., Markoff, L.: Specific requirements for elements of the 5′ and 3′ terminal regions in flavivirus RNA synthesis and viral replication. Virology **374**(1), 170–185 (2008)
5. Bollati, M., Alvarez, K., Assenberg, R., Baronti, C., Canard, B., Cook, S., Coutard, B., Decroly, E., de Lamballerie, X., Gould, E.A., Grard, G.: Structure and functionality in flavivirus NS-proteins: perspectives for drug design. Antivir. Res. **87**(2), 125–148 (2010)
6. Perera, R., Kuhn, R.J.: Structural proteomics of dengue virus. Curr. Opin. Microbiol. **11**(4), 369–377 (2008)
7. Brinton, M.A.: The molecular biology of West Nile Virus: a new invader of the western hemisphere. Ann. Rev. Microbiol. **56**(1), 371-402 (2002)
8. Jankowsky, E.: RNA helicases at work: binding and rearranging. Trends Biochem. Sci. **36**(1), 19–29 (2011)
9. Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D., Bairoch, A.: Protein identification and analysis tools on the ExPASy server. In: The Proteomics Protocols Handbook, pp. 571–607. Humana press (2005)
10. Avn, S.: Insilico exploration of PYC gene from corynebacterium glutamicum atcc 13032
11. Graña, O., Baker, D., MacCallum, R. M., Meiler, J., Punta, M., Rost, B., Tress, M.L., Valencia, A.: CASP6 assessment of contact prediction. Proteins Struct. Funct. Bioinf. **61**(S7), 214–224 (2005)
12. Liu, X.: Preparation and functional investigation of protease inhibitors conjugated to soluble dextran (2018)
13. McKee, L.S., Brumer, H.: Growth of Chitinophaga pinensis on plant cell wall glycans and characterisation of a glycoside hydrolase family 27 β-L-arabinopyranosidase implicated in arabinogalactan utilisation. PLoS ONE **10**(10), e0139932 (2015)
14. Bennett, M.J., Huey-Tubman, K.E., Herr, A.B., West, A.P., Ross, S.A., Bjorkman, P.J.: A linear lattice model for polyglutamine in CAG-expansion diseases. Proc. Natl. Acad. Sci. **99**(18), 11634–11639 (2002)
15. Vikram, T., Kumar, P.: Analysis of Hepatitis E virus (HEV) X-domain structural model. Bioinformation **14**(7), 398 (2018)
16. Bellingham, J., Foster, R.G.: Opsins and mammalian photoentrainment. Cell Tissue Res. **309**(1), 57–71 (2002)
17. Sandal, M., Duy, T.P., Cona, M., Zung, H., Carloni, P., Musiani, F., Giorgetti, A.: GOMoDo: a GPCRs online modeling and docking webserver. PLoS ONE **8**(9), e74092 (2013)