

Original Article

Characterizing the West Nile Virus's polyprotein from nucleotide sequence to protein structure – Computational tools

Mallari Praveen

Department of Zoology, Indira Gandhi National Tribal University, Amarkantak, Madhya Pradesh, India

Received 11 July 2023; revised 27 November 2023; accepted 8 January 2024; Available online 16 January 2024



المخلص

أهداف البحث: ينتمي فيروس غرب النيل إلى فصيلة الفيروسات المصفرة ويسبب حمى النيل الغربي، وآلية انتقاله تكون عن طريق نوع بعوض الكيولكس. الضحايا في الغالب لا تظهر عليهم أعراض، والقليل منهم يظهر عليهم الأعراض الشائعة، 10٪ من الخلايا العصبية التي يسببها هذا الفيروس سوف تسبب الوفاة. ومع ذلك، حتى الآن لا تزال البروتينات المشفرة بواسطة هذه الجينات غير مميزة، وفي هذا الصدد، فإن فهم وظائف وبنية هذه البروتينات له أهمية في صياغة الأدوية المضادة للفيروسات.

طريقة البحث: في الدراسة البحثية الحالية، تم استخدام منهجية "إن-سيليكو" مثل مجموعة متنوعة من أدوات المعلومات الحيوية وقواعد البيانات المستخدمة لتحليل بروتينات بروتينات فيروس غرب النيل بشكل فردي، والخصائص الفيزيائية والكيميائية، واكتشاف المجالات المحفوظة، وتحديد المناطق القابلة للذوبان وعبر الغشاء، وتوطين الإشارة، واضطراب البروتين وخصائص البنية الثانوية وهيكل البروتين ثلاثية الأبعاد الخاصة بكل منها.

النتائج: من بين أحد عشر بروتينا، ثمانية بروتينات أعلى من خمسين بالمائة من محتوى "جي سي"، وثمانية بروتينات ذات قيم "بي إل" الأساسية، وثلاثة بروتينات غير مستقرة في الظروف المختبرية، وأربعة بروتينات ذات قيمة "ايه أي" أعلى من 100 قابلة للحرارة، وبعض البروتينات ذات قيم "جرافي" سلبية من التحليل الفيزيائي والكيميائي. تتم مشاركة جميع المجالات المحفوظة بالبروتين مع عائلة الفيروسات المصفرة. خمسة بروتينات قابلة للذوبان ولا تحتوي على مناطق غشائية. يشير بروتينان إلى توطين الشبكة الإندوبلازمية المضيفة. أظهر غير الهيكلي 2 (إن إس) اضطرابا منخفضا في البروتين. تمثل السمات الهيكلية الثانوية وبيانات نماذج البنية الثالثة مصدرا كيميائيا حيويا قيما يوفر جزءا من المعلومات القوية لتصميم ركائز انتقائية ومثبطات اصطناعية.

الاستنتاجات: يمكن استخدام بروتينات فيروس غرب النيل "إن إس 2" و "إن إس 2ب" و "بي إم" و "إن إس 3" و "إن إس 5" كأهداف دوائية للدراسة الدوائية لتصميم مركبات الرصاص المضادة للفيروسات.

الكلمات المفتاحية: الأدوية المضادة للفيروسات؛ تصميم الأدوية؛ الفيروسات المصفرة؛ الجينوم؛ الفيزيوكيميائية؛ فيروس غرب النيل

Abstract

Objectives: West Nile virus (WNV) belongs to the Flaviviridae family and causes West Nile fever. The mechanism of transmission involves the culex mosquito species. Infected individuals are primarily asymptomatic, and few exhibit common symptoms. Moreover, 10 % of neuronal infection caused by this virus cause death. The proteins encoded by these genes had been uncharacterized, although understanding their function and structure is important for formulating antiviral drugs.

Methods: Herein, we used in silico approaches, including various bioinformatic tools and databases, to analyse the proteins from the WNV polyprotein individually. The characterization included GC content, physicochemical properties, conserved domains, soluble and transmembrane regions, signal localization, protein disorder, and secondary structure features and their respective 3D protein structures.

Results: Among 11 proteins, eight had >50 % GC content, eight proteins had basic pI values, three proteins were unstable under in vitro conditions, four were thermostable according to >100 AI values and some had negative GRAVY values in physicochemical analyses. All protein-conserved domains were shared among Flaviviridae family members. Five proteins were soluble and lacked transmembrane regions. Two proteins had signals for localization in the host endoplasmic reticulum. Non-

Corresponding address: Department of Zoology, Indira Gandhi National Tribal University, Amarkantak, Madhya Pradesh, India.

E-mail: mallaripraveen950@gmail.com

Peer review under responsibility of Taibah University.



Production and hosting by Elsevier

structural (NS) 2A showed low protein disorder. The secondary structural features and tertiary structure models provide a valuable biochemical resource for designing selective substrates and synthetic inhibitors.

Conclusions: WNV proteins NS2A, NS2B, PM, NS3 and NS5 can be used as drug targets for the pharmacological design of lead antiviral compounds.

Keywords: Antiviral drugs; Drug design; Flaviviridae; Genome; Physicochemical; West Nile Virus

© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

West Nile virus (WNV) is an insect-borne periodic epidemic disease causing seasonal infections in temperate climatic regions. This virus belongs to the genus *Flavivirus* of the Flaviviridae family.¹ Transmission occurs through culex species mosquitoes, which are exposed to the virus after feeding on dead birds.² Transmission through breastfeeding, blood transfusion and organ transplantation have been reported to lead to infections in humans.³ The first infection was reported in 1937, in Uganda. In the latter half of the 20th century, the virus spread worldwide.³ Most infected individuals are asymptomatic, and the few symptoms include general fever, vomiting, headache and rashes. Less than 1% of affected people exhibit neuroinvasive diseases, such as encephalitis and meningitis, 10% impact on the nervous system leading to death.

The genome of WNV is composed of a positive-stranded RNA containing approximately 10,000 nucleotides.⁴ Coding regions are translated into polyproteins, which are then cleaved into structural and non-structural proteins. The structural proteins are C, PM, M and E, whereas the non-structural (NS) proteins are NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5. The C (capsid) protein packs the RNA in an immature state.⁵ The PM and M proteins are considered a single entity known as M protein, which plays a crucial role in infection by activating the viral entry proteins within the cell.⁶ The envelope formed by the E protein binds the host cell's surface receptors.⁷ E protein is typically targeted by the T (CD8+) cell response.⁸ Replication complex regulation by NS1 imparts viral viability. Host cell death, viral assembly and replication are based on NS2A protein mechanisms.⁹

The cofactor NS2B plus NS3 form the NS2B-NS3 protease complex, which plays crucial roles in polyprotein cleavage and virion replication.⁵ NS3 has a multifunctional serine protease at the NTD and a helicase at the CTD.¹⁰ NS3 helicase activity is regulated by NS4A, and NS3 acts as a cofactor for infectious virion propagation.¹ NS4B blocks interferon signalling of the host cell. NS5, also known as RNA-dependent DNA polymerase, acts as a methyltransferase.¹¹

WNV protease (NS2B-NS3) activity is inhibited by 8-hydroxyquinoline¹² and the trypsin inhibitor aprotinin.¹³ Only the protease complex and E proteins have been used to design antiviral drugs by peptides and ligands. The remaining proteins have roles in replication and infection. Hence, detailed analysis of genome characteristics is a necessary prerequisite for developing drugs against WNV.

The present work focused on characterizing the 11,022 nucleotide WNV genome sequence retrieved from the NCBI database (<https://www.ncbi.nlm.nih.gov/nucleotide/11022>) by using various computational tools to assess nucleotide sequence and 3D protein structure (Figure 1). To our knowledge, no prior studies have performed complete genome characterization of proteins from the WNV polyprotein. Hence, this work may help medicinal chemists design new and better antiviral drugs on the basis of the described protein properties.

Materials and Methods

GC content calculation and translation

The polyprotein coding nucleotide sequence of WNV was retrieved from the NCBI database (<https://www.ncbi.nlm.nih.gov/>) (ID-KT862844.1). The nucleotide sequence encodes 11 proteins: C, PM, M, E, NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5. The GC content of the 11 protein nucleotide coding sequences was calculated with ENDMEMO (<http://www.endmemo.com/bio/gc.php>), and sequences were translated with EMBOSS TransSeq (https://www.ebi.ac.uk/Tools/st/emboss_transeq/)¹⁴ to obtain protein sequences with standard codons and without frame shifts.

Physicochemical properties of the proteins

The amino acid sequence of the proteins was input in ProtParam (<https://web.expasy.org/protparam/>) to analyse the physicochemical properties¹⁵; descriptors included the total number of amino acids, molecular weight, isoelectric constant (pI), total number of basic (Lys+Arg) and acidic (Glu+Asp) amino acids, extinction coefficient (EC),¹⁶ instability index (II),¹⁷ aliphatic index (AI)¹⁸ and grand average hydropathy (GRAVY).¹⁹

Conserved domain detection

NCBI's interface was used to search conserved domain detection (CDD) data for the 11 protein sequences (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>).²⁰ The deposited protein sequences were annotated with RPS-BLAST, a ψ -BLAST variant, to enhance the set of pre-calculated PPSSMs with query protein's, setting a threshold limit of 0.01.

Determination of soluble or transmembrane proteins

The SOUSI server (<https://hARRIER.nagahama-i-bio.ac.jp/sousi/mobile/>) was used to determine the soluble or conserved transmembrane residues of the proteins.²¹

Prediction of protein signal localization

The web tool Virus-mPLOC (<http://www.csbio.sjtu.edu.cn/bioinf/virus-multi/>), based on the data for 252 viral protein sequences, was used for predicting the viral protein's subcellular localization within the host and various sites of virus-infected cells.²²

Protein disorder

Under normal conditions, disordered proteins do not form well-defined tertiary structures. IUPred2A (<https://iupred2a.elte.hu/>),²³ a combined web interface, was used to determine protein disorder and disordered binding regions, by using IUPred2 and ANCHOR2.

Protein secondary structure features

Polyprotein secondary structure was determined with the web server Self Optimized Prediction Method and Alignment (SOPMA) (https://npsa-prabi.ibcp.fr/NPSA/npsa_sopma.html),²⁴ which provides detailed information on the α -helix, β -sheet, random coil and extended regions of proteins, given in fasta format.

Prediction of protein tertiary structures

Protein three-dimensional structure determination was performed through a template-based search with SWISS homology modelling (<https://swissmodel.expasy.org/interactive>).²⁵ ProMod3 was used for structural reading of input sequences along with insertions and deletions on an alignment basis; OpenMM and OpenStructure were used for simulations; and comparative modelling followed by parameterization was performed with the CHARMM22/CMAP force field. The automated models built structures based on the updated sequences from UniProtKB and 190,687 PDB repository structures, for template identification and determination of sequence identity. Generated tertiary structures were further subjected to validation in PROCHECK with the Saves v6.0 server,²⁶ to identify the percentage of residues in the allowed regions of the Ramachandran plots.

Abbreviations: 2D, Two- Dimensional/Structure; 3D, Three- Dimensional/Structure; CTD, Carboxy- Terminal Domain; EC, Extinction Coefficient; EMBOSS, European Molecular Biology Open Software Suite; ER, Endoplasmic Reticulum; GC, Guanine and Cytosine; GRAVY, Grand Average of Hydropathicity; II, Instability Index; NCBI, National Centre for Biotechnology Information; NS, Non-Structural; NTD, Nitrogen Terminal Domain; pI, Isoelectric Constant; PM, Pre-Membrane; PPSSM, Pseudo Position-Specific Scoring Matrix; RNA, Ribonucleic Acid; RPS-BLAST, Reverse Position Specific Basic Local Alignment Search Tool; WNV, West Nile Virus; QMEAN, Quantitative Model Analysis; RMSD, Root Mean Square Deviation.

Results

GC content calculation and translation

The total length of the WNV polyprotein was 11,022 nt; a 10,297 nucleotide coding sequence was found to encode a polyprotein (Table 1), which undergoes protease cleavage, thus forming the 11 proteins. NS5 had a nucleotide sequence length of 2715 nt, NS3 had a nucleotide sequence length of 1857 nt, M had a shorter nucleotide sequence length of 225 nt, and PM had a nucleotide sequence length of 275 nt. Among the 11 proteins, eight had >50%GC content: C, M, E, NS1, NS3, NS4A, NS4B and NS5. PM, NS2A and NS2B had 49.00–49.99 % GC content. NS4B had a high GC content of 52.15 %, and M had a GC content of 52.00 %. PM had the lowest GC content, at 49.09 %.

Physicochemical properties of the proteins

Each protein's physicochemical properties were computed with the ExpASY ProtParam tool. Descriptors including the length of the protein, EC, molecular weight, II, pI, AI, total number of positively (Arg+Lys) and negatively (Asp+Glu) charged residues, and GRAVY values were determined (Table 2). Amino acid composition, including polar, non-polar, aromatic, acidic and basic properties, are represented in Figures 2 and 3.

The pI values for these proteins ranged from 12.31 to 4.11. Four proteins (PM, NS1, NS2B and NS4A) had acidic pI values (not less than 3), seven proteins (C, M, E, NS2A, NS3, NS4B and NS5) were basic, and C protein had a pI value of 12.31. The number of negatively and positively charged residues was more in NS5, equal in PM, and lesser in M protein. EC values at 280 nm ranged from 12,615 to 2,22,730 M⁻¹ cm⁻¹. PM and NS4A shared the same low EC of 12,615 M⁻¹ cm⁻¹. The II values ranged from 51.77 (PM) to 25.97 (NS2B). PM, NS1 and NS2A had scores >40, thus indicating their instability. The remaining eight proteins were stable, with scores <40. The AI values ranged from 107.88 (NS4B) to 68.19 (PM). M, NS2A, NS4A and NS4B had AI values > 100. The GRAVY values for PM, NS1, NS3 and NS5 were negative, ranging from -0.185 (PM) to -0.583 (NS5), whereas the remaining proteins had values of 0.030 (C) to 0.790 (NS4A).

Conserved domain detection

In the CDD analysis of proteins from the WNV polyprotein, 4 of 11 proteins did not have domains identified from the database: PM, NS2A, NS2B and NS4B. Domains and descriptions for each protein are provided in Table 3. C, M and NS4A contained the CD of flavivirus family capsid protein, envelope glycoprotein M and the flavi_NS4 superfamily. E protein had two domain clans. Protein NS1 had a domain in the flavi_NS1 superfamily. NS3 contained three domain clans: flavivirus DEAD domain, viral

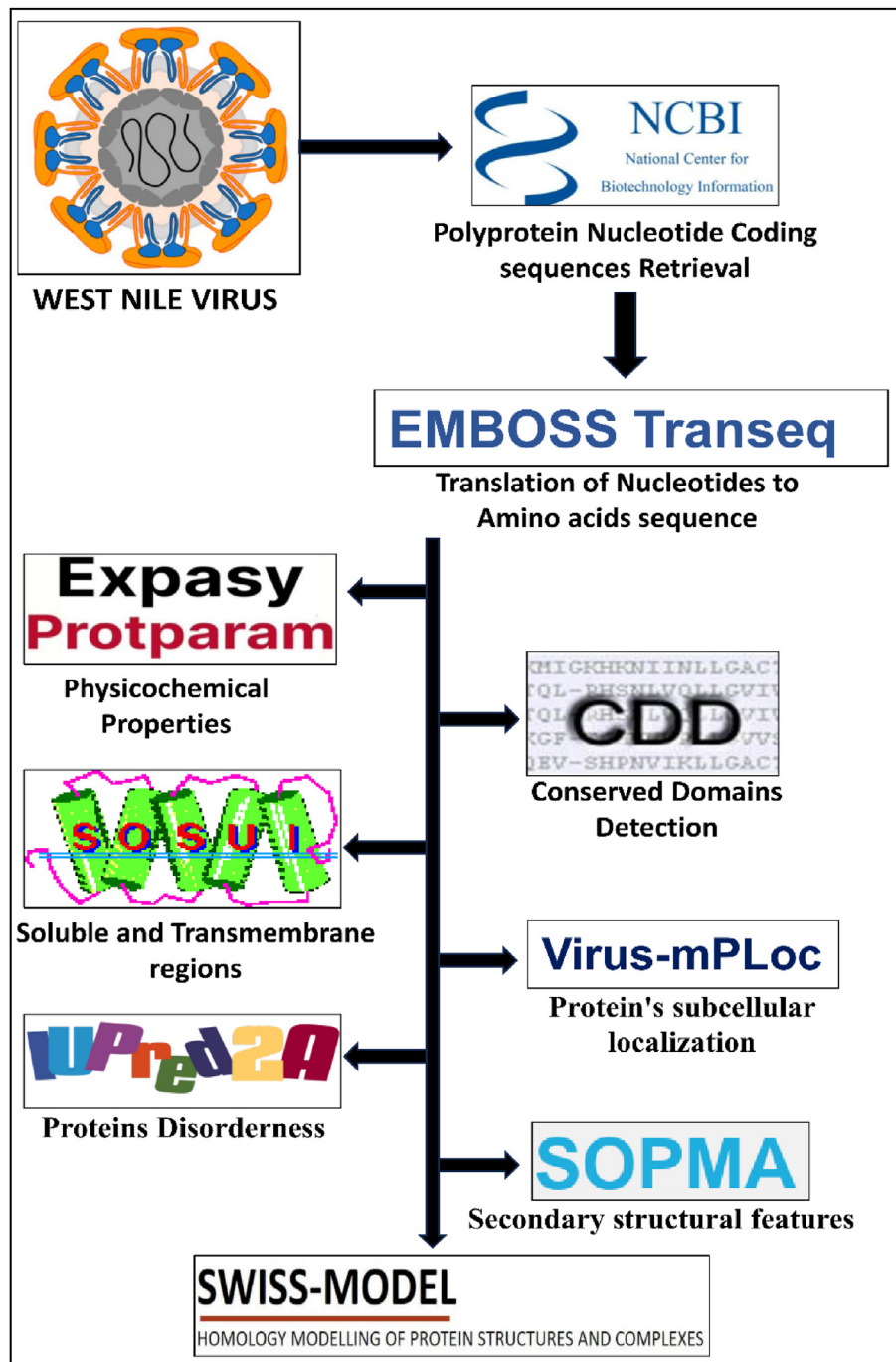


Figure 1: Methods used in the present study.

helicase CTD, peptidase S7 and flavivirus serine protease NS3. NS5 contained two domain clans: one relevant to the RNA-dependent DNA polymerase, which acts as the catalytic domain in the flavivirus genus, and one specific to the Flaviviridae methyltransferase.

Determination of soluble or transmembrane proteins

The SOUSI server was used to determine whether the proteins were soluble or transmembrane. M (2), NS2A (7),

NS2B (3), NS4A (4) and NS4B (4) contained transmembrane regions (Table 4 and Figure 4), whereas C, PM, NS1, NS3 and NS5 did not contain transmembrane regions and were soluble.

Prediction of protein signal localization

Investigation of the localization signals suggested that C, M, NS1, NS2A, NS2B, NS4A and NS4B localized in the viral capsid, whereas NS5 and NS3 localized in the viral

Table 1: Coding sequence length and GC content for proteins from the WNV polyprotein.

Nucleotides	Length	GC content (%)
C	369	51.49
PM	275	49.09
M	225	52.00
E	1503	51.29
NSP1	1056	50.75
NS2A	693	49.63
NS2B	392	49.23
NS3	1857	51.26
NS4A	447	51.67
NS4B	765	52.15
NS5	2715	51.82

C – Capsid; PM – Pre-Membrane; M – Membrane; E – Envelope; NS – Non-Structural Protein, GC – Guanine and Cytosine.

capsid and host endoplasmic reticulum, and PM protein localized in the host cytoplasm (Table 5).

Protein disorder

Protein region disorder, as described with IUPred2 (Figure 5), indicated binding region disorder with ANCHOR2. IUPred2 analysis indicated the following in descending order: PM>NS3>NS1>NS5>C>E>M>NS2B>NS4B>NS4A>NS2A. ANCHOR2 analysis indicated the following in descending order: NS3 >NS1 >NS5>PM>C>E>NS2B>M>NS4B>NS4A>NS2A.

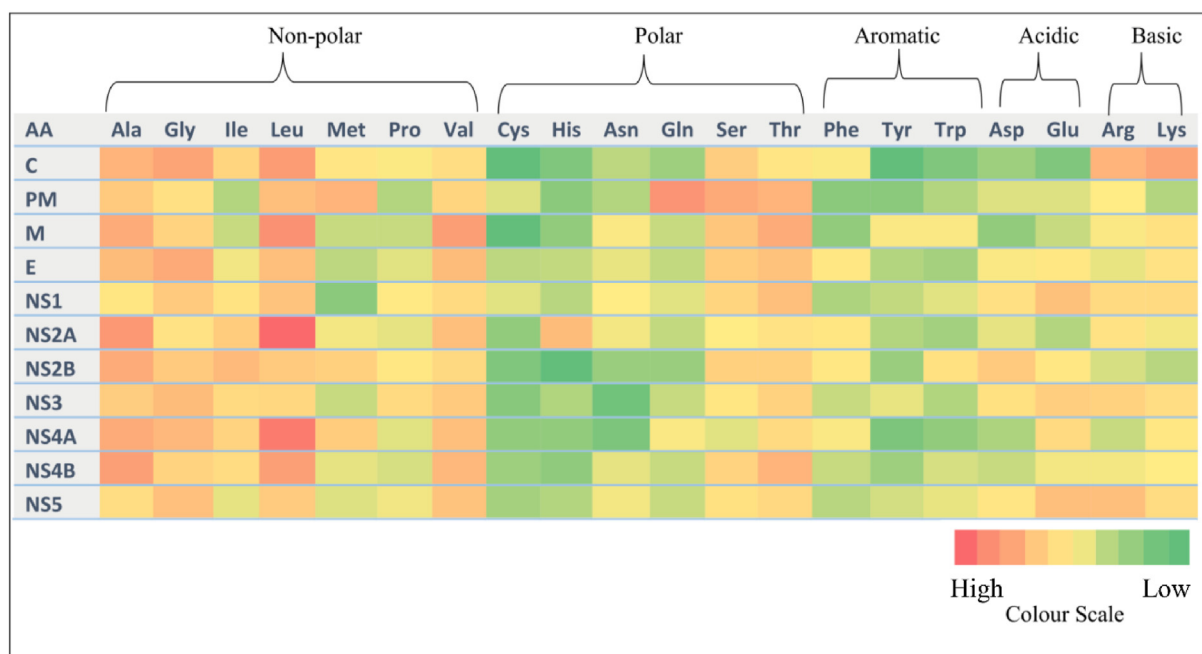
Protein secondary structure features

Secondary features of the proteins were predicted with the SOPMA tool (composition descriptors in Figure 6). The analysis revealed that the maximum frequency of

Table 2: Physico-chemical properties of proteins from the WNV polyprotein.

Sl. No.	Protein	AA	M.wt.	pI	(-)R	(+)R	EC ^a	II	AI	GRAVY
1	C	123	13315.15	12.31	3	26	5500	37.66 (S)	99.19	0.030
2	PM	92	10191.73	6.72	6	6	12,615	51.77(US)	68.19	-0.185
3	M	75	8195.58	9.87	3	7	20,970	36.64(S)	107.87	0.311
4	E	501	53592.29	7.66	43	44	66,640	24.10(S)	84.07	0.053
5	NS1	352	39731.77	5.73	49	42	80,160	45.87(US)	74.20	-0.568
6	NS2A	231	25391.57	9.69	13	21	29,575	42.24(US)	133.90	0.790
7	NS2B	131	14469.98	4.11	16	7	41,480	25.97(S)	93.89	0.390
8	NS3	619	69051.48	7.32	77	77	104,655	31.28(S)	77.17	-0.449
9	NS4A	149	16062.41	6.08	12	11	12,615	33.14(S)	122.42	0.760
10	NS4B	255	27546.28	8.93	17	21	50,210	30.60(S)	107.88	0.419
11	NS5	905	103622.92	8.56	124	132	222,730	36.79(S)	72.87	-0.583

AA – Total Number of Amino Acids; M.wt – Molecular Weight; pI – Isoelectric Constant; (-)R – Total Number of Negatively Charged Residues (Asp+Glu); (+)R – Total Number of Positively Charged Residues (Arg+Lys); EC – Extinction Coefficient (^a units of M⁻¹ cm⁻¹ at 280 nm measured in water); II – Instability Index; AI – Aliphatic Index; GRAVY – Grand Average of Hydropathicity

**Figure 2: Heat map representation of amino acid composition in proteins from the WNV polyprotein.**

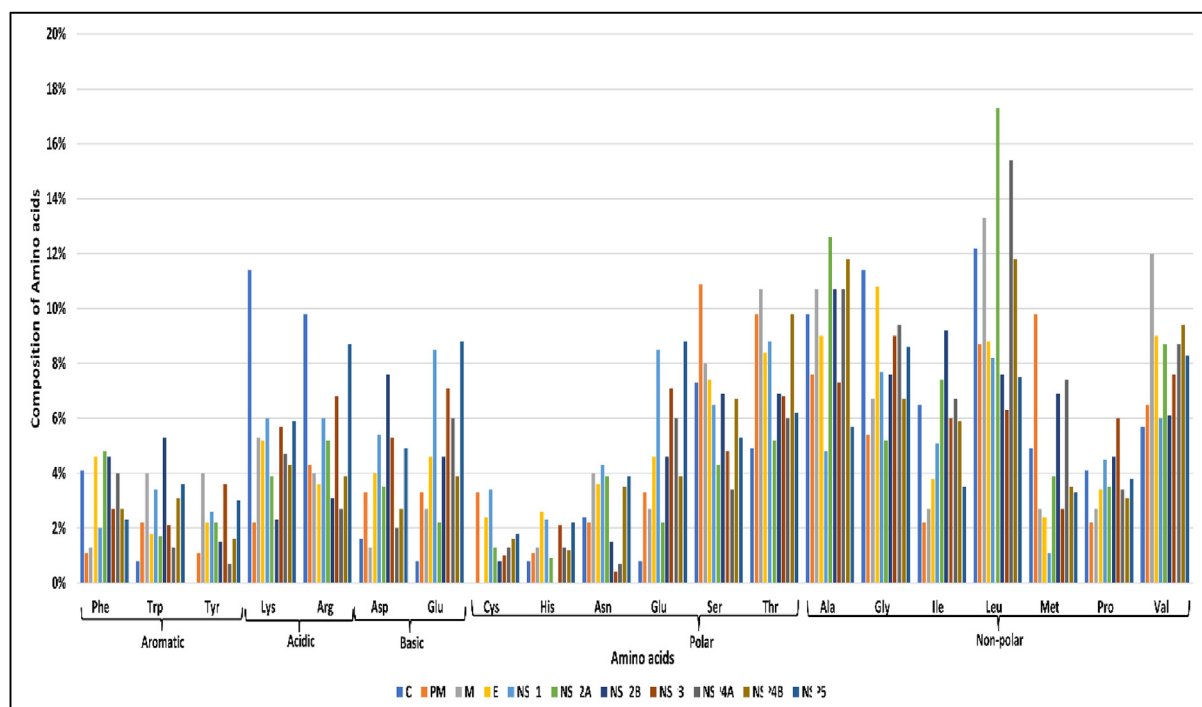


Figure 3: Amino acid residue composition of proteins from the WNV polyprotein.

Table 3: Identified conserved domains and descriptions for proteins from the WNV polyprotein.

Protein	Domain	Description
C	Flavi_capsid	Flavivirus capsid protein C
PM	—	—
M	Flavi_M superfamily	Flavivirus envelope glycoprotein M
E	Flavi_E_C Flavi_E_stem	Immunoglobulin-like domain III of Flavivirus envelope glycoprotein E Flavivirus envelope glycoprotein E, stem/anchor domain
NS1	Flavi_NS1 super family	Flavivirus NS1
NS2A	—	—
NS2B	—	—
NS3	Flavi_DEAD SF2_C_viral Peptidase_S7	Flavivirus DEAD domain Viral helicase CTD Flavivirus NS3 serine protease
NS4A	Flavi_NS4A superfamily	Flavivirus NS4A encoded as a single polyprotein
NS4B	—	—
NS5	Flavivirus_RdRp capping_2-OMTase Flaviviridae	RdRp catalytic core domain in Flavivirus genus. Cap-0 specific (nucleoside-2'-O-)-methyltransferase of Flaviviridae

C – Capsid; PM – Pre-Membrane; M – Membrane; E – Envelope; NS – Non-Structural Protein.

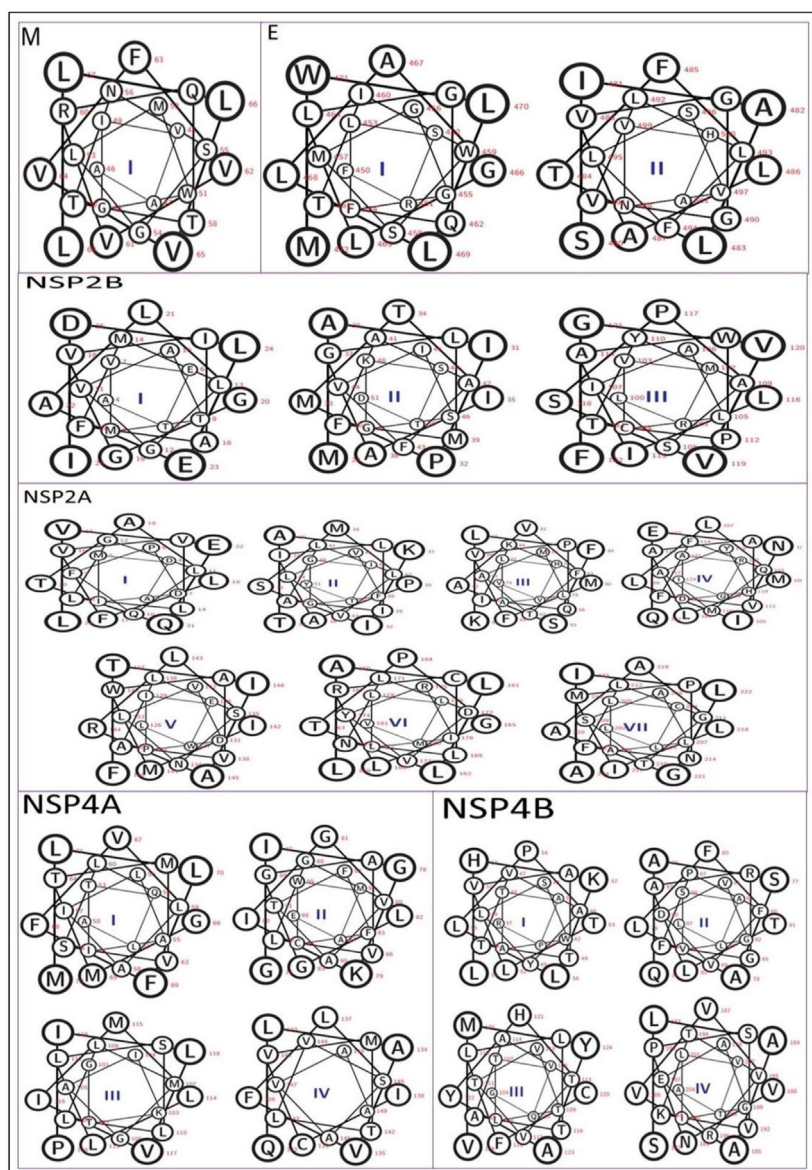
Table 4: Functional characterization of proteins from the WNV polyprotein.

Protein	S/M	TM regions				Type of Peptide sequence
		NTD	Sequence	CTD	Length	
C	S	—	—	—	—	—
PM	S	—	—	—	—	—
M	M	46	AAVIGWMLGSNTTQRVVFVLL	68	23	Primary
E	M	450	FRSLFGGMSWITQGLLGALLLWM	472	23	Secondary
		480	SIALTFLAVGGVLLFSLVNVHA	501	22	Primary
NS1	S	—	—	—	—	—
NS2A	M	3	ADMIDPFQLGLLVVFLATQEV	24	22	Secondary
		29	TAKISMPAILIALLVLFVGGITY	51	23	Primary
		74	VVHLALMATFKIQPFVFMVASFLK	96	23	Secondary
		102	QENILLMLAAVFFQVMAYHDARQI	124	23	Secondary
		126	LWEIPDVLNSLAVAWMILRAITF	148	23	Primary
		159	LALLTPGLRCLNLDVYRILLMV	181	23	Secondary
		202	LLCLALASTGLFNPMLAAGLIA	224	23	Secondary

(continued on next page)

Table 4 (continued)

Protein	S/M	TM regions				Type of Peptide sequence
		NTD	Sequence	CTD	Length	
NS2B	M	4	ATEVMTAVGLMFAIVGGLAELDI	26	23	Primary
		29	MAIPMTIAGLMFAAFVISGKSTD	51	23	Primary
		100	LRMFLCAISAYTPWAILPSVVG	122	23	Primary
NS3	S	—	—	—	—	—
NS4A	M	50	ALQTIALIALLSVMTMGVFFLLM	72	23	Primary
		76	GIGKIGLGGAVLGVATFFCWMAE	98	23	Primary
		101	GTKIAGMLLLSLLMIVLIP	120	19	Primary
		132	QLAVFLICVMTLVSAVAA	149	17	Primary
NS4B	M	37	RPATAWSLYAVTTAVLTPLLKHL	59	23	Secondary
		75	QASALFTLARGFPFVDVGVSA	97	23	Secondary
		104	GQVTLTVTVTAATLLFCHYAYMV	126	23	Primary
		182	SLAAVVVNPSVKTVREAGILVTA	204	23	Secondary
NS5	S	—	—	—	—	—



C – Capsid; PM – Pre-Membrane; M – Membrane; E – Envelope; NSP – Non-Structural Protein

Figure 4: Representation of transmembrane regions in wheel form.

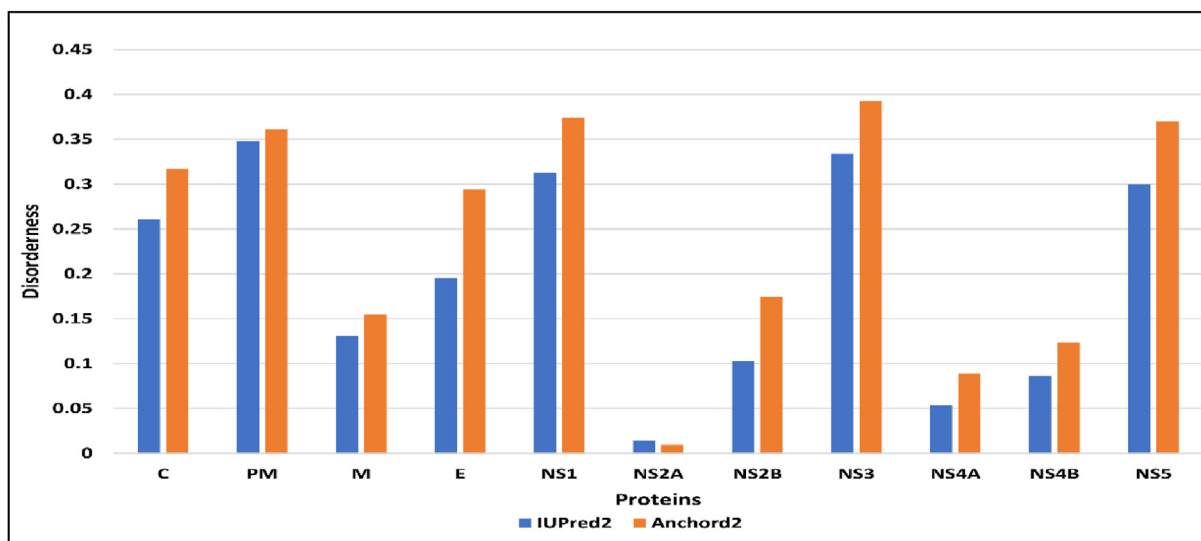
Table 5: Subcellular localization of the proteins from the WNV polyprotein.

Protein	Subcellular location
C	Viral capsid
PM	Host Cytoplasm
M	Viral capsid
E	Viral capsid
NS1	Viral capsid
NS2A	Viral capsid
NS2B	Viral capsid
NS3	Viral capsid, host ER
NS4A	Viral capsid
NS4B	Viral capsid
NS5	Viral capsid, host ER

occurrence of α -helices was in NS2A (67.53), and the minimum was in E (21.56). Relatively more random coils were present in NS1 (46.88), whereas fewer were present in C (5.69). Relatively more extended turns were present in NS3 (23.42), whereas fewer were present in NS2A (7.36). The β -sheet content was relatively low in M (4.00) and high in NS2B (9.16).

Prediction of protein tertiary structures

The tertiary structures of the proteins from the WNV polyprotein models were matched to the structures deposited in the PDB (Table 6), including the percentage sequence identity and alignment sequences. The templates used were 5OW2 for C, 2BSJ for PM, 7 LCG for M, 7KV9 for E,

**Figure 5:** Disorder of proteins from the WNV polyprotein.

IUPred2 – Disorder score of the protein region; Ancho2 – Disorder score of the binding region

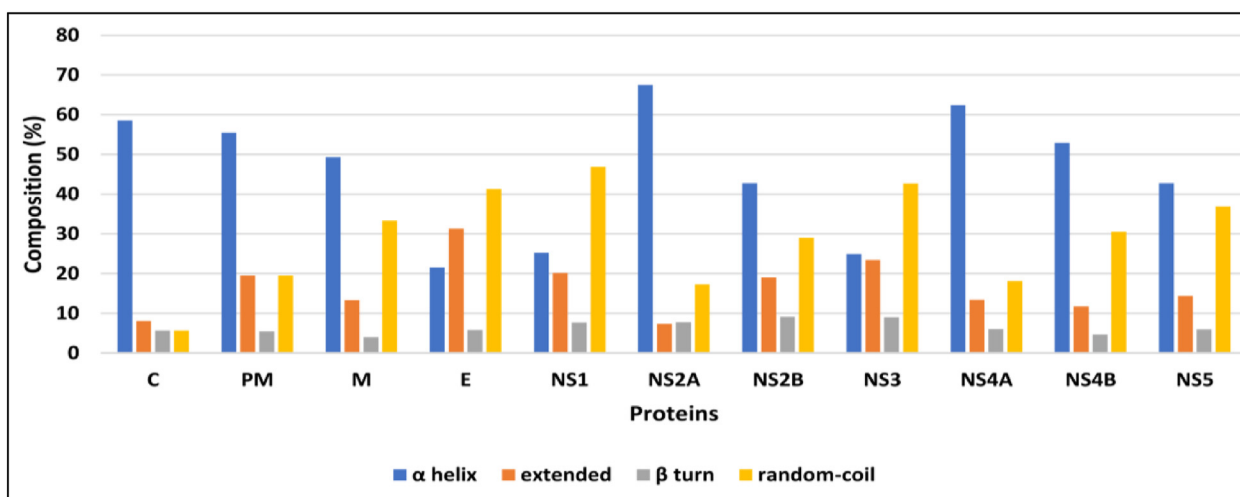
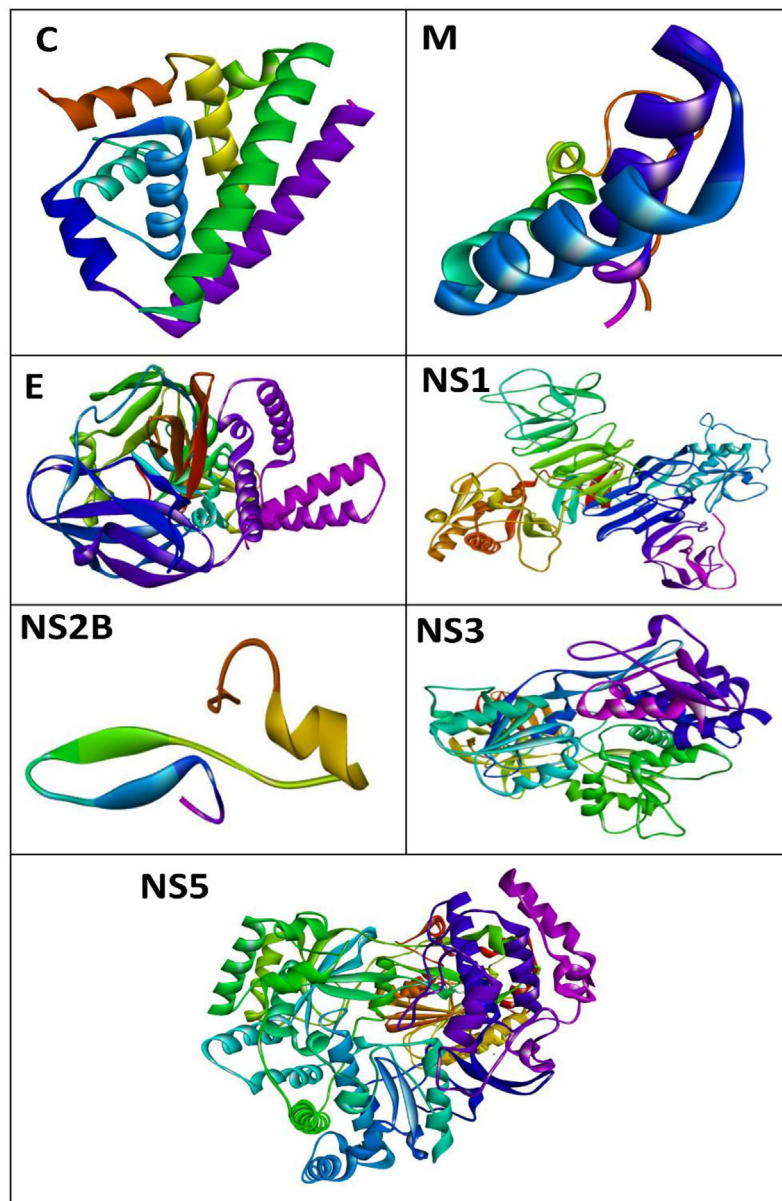
**Figure 6:** Secondary structure features of proteins from the WNV polyprotein.

Table 6: Shared protein templates from the PDB with respect to proteins from the WNV polyprotein and validation in Ramachandran plots.

Protein	Identity (%)	PDB ID	Chain	Molecule	Alignment	Ramachandran plot allowed regions (%)
C	66.35	5OW2	AB	Japanese encephalitis capsid protein	26–98	98.4
PM	23.68	2BSJ	AB	Type 3 secretion chaperone SycT from <i>Yersinia enterocolitica</i>	7–44	92.9
M	80	7LCG	F	Mature Usutu SAAR-1776	1–75	90.9
E	97.01	7KV9	A	Chimeric flavivirus between Binjari and WNV	1–501	92.5
NS1	99.72	4TPL	AB	WNV ns1	1–352	88.4
NS2A	13.11	6ZLH	A	Glutamate receptor transporter homologue Glt Tk	51–121	86.2
NS2B	93.75	2GGV	A	WNV Serine protease subunit ns2b-ns3	50–93	94.7
NS3	79.42	2WV9	A	Ns3 protease-helicase from Murray Valley encephalitis virus	16–618	90.6
NS4A	17.72	6HUM	F	Photosynthetic complex from <i>Thermosynechococcus elongatus</i>	46–124	92.2
NS4B	16.33	5H3E	B	Isocitrate dehydrogenase 2	104–161	90.2
NS5	81.19	4K6M	A	Japanese encephalitis virus ns5	5–895	94.4

C – Capsid; PM – Pre-Membrane; M – Membrane; E – Envelope; NS – Non-Structural Protein.

**Figure 7:** Predicted 3D structures of proteins from the WNV polyprotein.

4TPL for NS1, 6ZLH for NS2A, 2QQV for NS2B, NS2B for NS3, 6HUM for NS4A, 5H3C for NS4B and 4K6M for NS5. A sequence identity >30 % was used to build the 3D protein structure. The built predicted 3D protein structures for C, M, E, NS1, NS2B, NS3 and NS5 are shown in [Figure 7](#). Validation of the generated 3D protein structures with Ramachandran plots was based on >90 % of residues in allowed regions indicating the built model was appropriate. Only two proteins, NSP1 and NS2A, had <90 % of residues in allowed regions ([Table 6](#) and [Figs. S1–S11](#)).

Discussion

The guanine (G) and cytosine (C) content in the nucleotide sequence affects DNA stability. A high GC content enhances genome stability at elevated temperatures and promotes strong base stacking.²⁷ It imparts crucial information for determining the temperature during PCR annealing, designing primers, and constitutes a key feature in shaping proteins.²⁸ The proteins NS4B, M, NS5, NS4A, C, E, NS3 and NS1 were found to have high GC content of 52.15 %, 52.00 %, 51.82 %, 51.67 %, 51.49 %, 51.29 %, 51.26 % and 50.75 %, respectively, accounting for more than 50 % of the genome; consequently, the proteins that are stable and rigid, and can withstand relatively higher temperatures than other WNV polyproteins. In PCR, when generating complementary gene sequences for these proteins, the annealing temperature is slightly higher for proteins with lower GC composition. Proteins with less stability and greater flexibility are more prone to mutations under abnormal conditions.

Physicochemical analysis was performed through an *in silico* approach. The pI is the pH at which no electrical charge is present on the molecule, or the total number of negative and positive charges is equal.¹⁵ Four proteins (PM, NS1, NS2B and NS4A) had pI values below seven, whereas the remaining seven proteins had pI values above seven. The isoelectric focusing technique is performed on the basis of the pI values to separate molecules from complexes.²⁹ These values aid in the isolation of proteins of interest from the WNV polyprotein in wet laboratory experiments after digestion.

EC is defined as the amount of light absorbed per mole protein at a specific wavelength of light. A protein's EC value is calculated according to composition of tryptophan, tyrosine, and cysteine residues, because these amino acids substantially contribute to measuring the protein's optical density in the 276–282 nm range.¹⁵ Protein–protein and protein–ligand quantitative studies can be understood on the basis of EC values.³⁰ The highest EC value was observed for the NS5 protein. II indicates protein stability under both *in vivo* and *in vitro* conditions. Proteins with II above 40 are considered unstable, whereas those with II below 40 are stable.¹⁷ The proteins PM, NS1 and NS2A were found to be unstable; thus, efficient procedures are required for laboratory experiments. The remaining eight proteins were found to be stable. AI is another parameter describing protein stability according to temperature. AI is defined as the relative volume occupied by aliphatic side chains, such

as alanine (Ala), valine (Val), leucine (Leu) and isoleucine (Ile).^{15,18} A high AI indicates high thermostability of a protein, which is an additional factor for wet laboratory studies. Among proteins from the WNV polyprotein, NS2A (133.90) and NS4A (122.42) had high AI values indicating high stability under a wide range of temperatures. IN contrast, PM (68.15) and NS5 (72.87) had low AI values and highly flexible structures. Proteins with lower GC and AI values were more flexible, as indicated in [Tables 1 and 2](#). GRAVY values range from –4 to +4, and indicate the hydrophilic and hydrophobic nature of proteins.¹⁹ A low GRAVY range suggests a globular (hydrophobic) rather than a membranous (hydrophilic) protein. In this study, the proteins with negative GRAVY values (PM, E, NS3 and NS5) ([Table 2](#)) did not form transmembrane regions, as confirmed by the SOSUI server ([Table 3](#)).

A set of proteins from the WNV polyprotein subjected to NCBI/CDD-BLAST is listed in [Table 3](#). The proteins were found to contain conserved domains, such as Flavi_capsid, Flavi_M superfamily, Flavi_E_C, Flavi_E_Stem, Flavi_NS1 superfamily, Flavi_DEAD, SF_2_C_viral, pepridase_S7, Flavi_NS4A superfamily, Flavi_RdRp and capping_2_OMTase-Flaviviridae. Few of these domains have enzymatic activity. Four proteins, PM, NS2A, NS2B and NS4B, were found not to have conserved domains, according to the CDD search. A superfamily is a set of conserved domain models that generate overlapping annotations of protein sequences, are assumed to represent evolutionarily related domains and may be redundant with one another.³¹ The superfamily is relatively close to being identified as the Flaviviridae superfamily.

Four amino acid characteristics are applied in SOSUI predictions: hydrophathy index, amphiphilicity index, amino acid sequence charge and protein length.²¹ The bilayer transmembrane functions in physiological process such as cell recognition, intracellular joining, attachment, enzymatic function and signal transduction.³² The results in [Table 4](#) indicated that proteins with TM regions contained high non-polar amino acids to maintain the hydrophobicity and relatively less polar residue content. NS2A, NS4A, NS4B, NS2B, E and M contain TM regions that actively perform cellular functions.

Viruses can replicate their genome to increase their progeny only in host cells, and their functions depend on the environmental conditions in the body.³³ Therefore, knowledge of the subcellular localization of viral proteins in host cells or virus-infected cells greatly aids in understanding the relevant mechanisms and designing antiviral drugs. On the basis of analysis of the subcellular localization of WNV proteins ([Table 5](#)), all proteins except PM were predicted to localize in the viral capsid. NS3 and NS5 were predicted to localize in the host endoplasmic reticulum, and PM protein was predicted to localize in the host cytoplasm. The translation of WNV RNA in the host ER results in polyprotein formation, increases ER stress and leads to protein unfolding.³⁴

Disordered proteins have relatively low hydrophobic residue content and high amounts of charged polar residues, the latter of which are responsible for hydrophilicity and consequently interference with water molecules and the

formation of weak multivariant interactions.^{35,36} Analysis of WNV polyprotein disorder (Figure 5) indicated that NS2A was less disordered and more stable than NS3, in terms of both protein regions and binding region disorder.

Among all proteins, the β -sheet composition was greatest in NS2B (Figure 6). Graces et al., in X-ray crystallography studies, have observed that NS2B comprises more β -sheets,³⁷ in agreement with the results of this study. Tertiary structure has broad applications in the pharmacologic design of drugs targeting a given protein. The tertiary structures of proteins from the WNV polyprotein are represented in Figure 7. However, protein structure refinement and analysis through modelling must be performed to use these structures accurately. Nonetheless, our validation results indicated that most of the generated proteins had residues in allowed regions in the Ramachandran plot. Further structural refinements, such as QMEAN scores and RMSD, have been suggested to extend the study of these proteins. The proteins PM, NS2A, NS4A and NS4B showed <30 % identity (Table 6), and ab initio modelling must be performed to obtain their 3D structures.

In current study, we recognize the limitation of potential sampling bias, wherein the chosen proteins might not have fully captured WNV polyprotein diversity. This bias might affect the external validity or generalizability of our findings to the broader WNV population. To mitigate this limitation, future research could explore a more diverse set of proteins or consider comparisons with other WNV strains or Flaviviridae family viruses. Although this study provides valuable understanding within the scope of the selected proteins, we acknowledge the importance of expanding the analysis to gain a more comprehensive understanding.

The methods used in the current study were formulated on the basis of insights from previously published research articles.^{38,39} Although every bioinformatics tool has its own advantages, being aware of limitations is crucial. For example, working with very large datasets may lead to difficulties with EMBOSS TransSeq, thus delaying processing times. ProtParam is a useful program for calculating basic protein parameters but might be unable to capture finer details, particularly in the case of sophisticated post-translational modifications or interactions with other biomolecules. Because NCBI-CDD is dependent on pre-existing databases, its accuracy relies on the comprehensiveness of those databases. Whereas SOUSI, Virus-mPLoc and IUPRED2 provide useful information about subcellular localization and protein abnormalities, they might not be able to accurately anticipate fine structural features or proteins with significant divergence. Furthermore, ANCHOR2 is useful in locating disordered binding areas, but care must be taken, because its predictions might not always align perfectly with experimental observations. SOPMA can predict secondary structures, but its accuracy varies, particularly in areas with large structural differences. SWISS-Homology modelling is a powerful tool whose efficacy is contingent on the quality and accessibility of template structures in the Protein Data Bank. Because proteins are dynamic, and computational predictions have inherent limits, results must

be evaluated cautiously. These bioinformatics tools are best used in conjunction with experimental validation to support strong conclusions.

The conserved domain characteristics of WNV, in comparison to those of other flaviviruses, are highly conserved across the genus and thus may serve as potential targets for novel therapeutic strategies.⁴⁰ A large-scale analysis of the WNV proteome has revealed the numerous evolutionarily stable nonameric positions present across the proteome and identified several completely conserved sequences.⁴¹ Mutational studies on the fusion glycoproteins indicated they inhibited the Zika virus and yellow fever virus, but not WNV, in terms of production of infectious virions. These conserved sequences are shared by other flaviviruses and have been associated with the functional and structural properties of viral proteins.⁴² Additionally, the WNV envelope glycoprotein fusion peptide region has been identified as an immunodominant epitope stimulating antibodies, particularly monoclonal antibodies with diverse patterns of cross-reactivity.⁴³ Understanding these conserved domain characteristics is important for the development of antiviral therapies and the design of peptide-specific vaccines for flavivirus infections.⁴⁴

Because in silico methods were used, additional study limitations include the valid lack of wet laboratory (in vivo and in vitro) investigation. The computational tools and servers, after updating or reconstruction with various algorithms, might yield different results even if the same sequence of the WNV is input. Future investigations based on the key findings of this study might consider the non-structural and pre-membrane proteins of the WNV polyprotein. Notably, NS2A, NS2B, PM, NS3 and NS5 proteins may be favourable drug targets for designing small molecules/ligands in drug discovery⁴⁵ and epitope design through immunoinformatic approaches.⁴⁶

Conclusion

The proteins derived from the WNV polyprotein are associated with viral replication and induce disease in host cells. In this present study, several bioinformatic tools were used to study the characteristics of the WNV genome, including basic nucleotide sequences and complex 3D protein structures for structural and non-structural proteins. NS4B protein, with its high GC content, is considered relatively stable. In physicochemical analysis, the pI of C protein was highly basic, whereas that of NS2B was weakly acidic. The total number of positive and negative amino acid residues was relatively greater in NS5, thus indicating its high reactivity and ability to be isolated easily from the complex, given its high EC value. NS2A was the most thermostable among all proteins, on the basis of its high AI. Four proteins did not have transmembrane domains, on the basis of negative GRAVY values. Conserved domains were identified from the Flaviviridae family. The SOSUI server identified seven proteins with transmembrane domains. The 2D and 3D structures of these proteins provide insights for developing accurate models. Therefore, these proteins might be

crucial for host invasion and therefore could potentially be used as drug targets for pharmacological study.

Source of funding

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Conflict of interest

The author declares that there are no conflicts of interest regarding the publication of this article.

Ethical approval

There are no ethical issues.

Authors contributions

The author has critically reviewed and approved the final draft, and is responsible for the content and similarity index of the manuscript.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author used ChatGPT and QuillBOT to improve the language. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jtumed.2024.01.001>.

References

- Colpitts TM, Conway MJ, Montgomery RR, Fikrig E. West Nile Virus: biology, transmission, and human infection. *Clin Microbiol Rev* 2012; 25(4): 635–648. <https://doi.org/10.1128/CMR.00045-12>. PMID: 23034323; PMCID: PMC3485754.
- Mackenzie JS, Gubler DJ, Petersen LR. Emerging flaviviruses: the spread and resurgence of Japanese encephalitis, West Nile and dengue viruses. *Nat Med* 2004; 10: S98–S109. <https://doi.org/10.1038/nm1144>. PMID: 15577938.
- Sampathkumar P. West Nile virus: epidemiology, clinical presentation, diagnosis, and prevention. *Mayo Clin Proc* 2003; 78(9): 1137–1143. <https://doi.org/10.4065/78.9.1137>. quiz 1144. PMID: 12962168; PMCID: PMC7125680.
- Mukhopadhyay S, Kim BS, Chipman PR, Rossmann MG, Kuhn RJ. Structure of West Nile virus. *Science* 2003; 302(5643): 248. <https://doi.org/10.1126/science.1089316>. PMID: 14551429.
- Londono-Renteria B, Colpitts TM. A brief review of West Nile virus biology. *Methods Mol Biol* 2016; 1435: 1–13. https://doi.org/10.1007/978-1-4939-3670-0_1. PMID: 27188545.
- Moesker B, Rodenhuis-Zybert IA, Meijerhof T, Wilschut J, Smit JM. Characterization of the functional requirements of West Nile virus membrane fusion. *J Gen Virol* 2010; 91(2): 389–393. <https://doi.org/10.1099/vir.0.015255-0>. PMID: 19828760.
- Perera-Lecoin M, Meertens L, Carnec X, Amara A. Flavivirus entry receptors: an update. *Viruses* 2013; 6(1): 69–88. <https://doi.org/10.3390/v6010069>. PMID: 24381034; PMCID: PMC3917432.
- Larsen MV, Lelic A, Parsons R, Nielsen M, Hoof I, Lamberth K, Loeb MB, Buus S, Bramson J, Lund O. Identification of CD8+ T cell epitopes in the West Nile virus polyprotein by reverse-immunology using NetCTL. *PLoS One* 2010; 5(9): e12697. <https://doi.org/10.1371/journal.pone.0012697>. PMID: 20856867; PMCID: PMC2939062.
- Youn S, Ambrose RL, Mackenzie JM, Diamond MS. Non-structural protein-1 is required for West Nile virus replication complex formation and viral RNA synthesis. *Virol J* 2013; 10: 339. <https://doi.org/10.1186/1743-422X-10-339>. PMID: 24245822; PMCID: PMC3842638.
- Shiryaev SA, Strongin AY. Structural and functional parameters of the flaviviral protease: a promising antiviral drug target. *Future Virol* 2010; 5(5): 593–606. <https://doi.org/10.2217/fvl.10.39>. PMID: 21076642; PMCID: PMC2976050.
- Davidson AD. Chapter 2. New insights into flavivirus nonstructural protein 5. *Adv Virus Res* 2009; 74: 41–101. [https://doi.org/10.1016/S0065-3527\(09\)74002-3](https://doi.org/10.1016/S0065-3527(09)74002-3). PMID: 19698895.
- Capone G, Pagoni M, Delfino AP, Kanduc D. Evidence for a vast peptide overlap between West Nile virus and human proteomes. *J Basic Microbiol* 2013; 53(10): 800–807. <https://doi.org/10.1002/jobm.201200204>. PMID: 22961336.
- Mueller NH, Yon C, Ganesh VK, Padmanabhan R. Characterization of the West Nile virus protease substrate specificity and inhibitors. *Int J Biochem Cell Biol* 2007; 39(3): 606–614. <https://doi.org/10.1016/j.biocel.2006.10.025>. PMID: 17188926.
- Madeira F, Pearce M, Tivey ARN, et al. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res* 2022; gkac240. <https://doi.org/10.1093/nar/gkac240>. PMID: 35412617; PMCID: PMC9252731.
- Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF. Protein identification and analysis tools in the ExpASY server. *Methods Mol Biol* 1999; 112: 531–552. <https://doi.org/10.1385/1-59259-584-7:531>. PMID: 10027275.
- Gill SC, von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* 1989 Nov 1; 182(2): 319–326. [https://doi.org/10.1016/0003-2697\(89\)90602-7](https://doi.org/10.1016/0003-2697(89)90602-7). Erratum in: *Anal Biochem* 1990; 189(2): 283. PMID: 2610349.
- Guruprasad K, Reddy BV, 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. *Protein Eng* 1990; 4(2): 155–161. <https://doi.org/10.1093/protein/4.2.155>. PMID: 2075190.
- Ikaï A. Thermostability and aliphatic index of globular proteins. *J Biochem* 1980; 88(6): 1895–1898. PMID: 7462208.
- Kyte J, Doolittle RF. A simple method for displaying the hydrophobic character of a protein. *J Mol Biol* 1982; 157(1): 105–132. [https://doi.org/10.1016/0022-2836\(82\)90515-0](https://doi.org/10.1016/0022-2836(82)90515-0). PMID: 7108955.
- Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokkandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* 2011 Jan; 39(Database issue): D225–D229. <https://doi.org/10.1093/nar/gkq1189>. PMID: 21109532; PMCID: PMC3013737.
- Hirokawa T, Boon-Chieng S, Mitaku S. SOSUI: classification and secondary structure prediction system for membrane

- proteins. **Bioinformatics** 1998; 14(4): 378–379. <https://doi.org/10.1093/bioinformatics/14.4.378>. PMID: 9632836.
22. Shen HB, Chou KC. Virus-mPLoc: a fusion classifier for viral protein subcellular location prediction by incorporating multiple sites. **J Biomol Struct Dyn** 2010; 28(2): 175–186. <https://doi.org/10.1080/07391102.2010.10507351>. PMID: 20645651.
 23. Erdős G, Dosztányi Z. Analyzing protein disorder with IUPred2A. **Curr Protoc Bioinformatics** 2020; 70(1): e99. <https://doi.org/10.1002/cpbi.99>. PMID: 32237272.
 24. Geourjon C, Deléage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. **Comput Appl Biosci** 1995; 11(6): 681–684. <https://doi.org/10.1093/bioinformatics/11.6.681>. PMID: 8808585.
 25. Studer G, Tauriello G, Bienert S, Biasini M, Johner N, Schwede T. ProMod3-A versatile homology modelling toolbox. **PLoS Comput Biol** 2021; 17(1): e1008667. <https://doi.org/10.1371/journal.pcbi.1008667>. PMID: 33507980; PMCID: PMC7872268.
 26. Laskowski RA, MacArthur MW, Thornton JM. PROCHECK: validation of protein structure coordinates. In: Rossmann MG, Arnold E, editors. *International tables of crystallography, volume F. Crystallography of biological macromolecules*. The Netherlands: Dordrecht, Kluwer Academic Publishers; 2001. pp. 722–725.
 27. Yakovchuk P, Protozanova E, Frank-Kamenetskii MD. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. **Nucleic Acids Res** 2006; 34(2): 564–574. <https://doi.org/10.1093/nar/gkj454>. PMID: 16449200; PMCID: PMC1360284.
 28. Pozzoli U, Menozzi G, Fumagalli M, Cereda M, Comi GP, Cagliani R, Bresolin N, Sironi M. Both selective and neutral processes drive GC content evolution in the human genome. **BMC Evol Biol** 2008; 8: 99. <https://doi.org/10.1186/1471-2148-8-99>. PMID: 18371205; PMCID: PMC2292697.
 29. Pergande MR, Cologna SM. Isoelectric point separations of peptides and proteins. **Proteomes** 2017; 5(1): 4. <https://doi.org/10.3390/proteomes5010004>.
 30. Mohan R, Venugopal S. Computational structural and functional analysis of hypothetical proteins of *Staphylococcus aureus*. **Bioinformation** 2012; 8(15): 722–728. <https://doi.org/10.6026/97320630008722>. PMID: 23055618; PMCID: PMC3449381.
 31. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. The Pfam protein families database: towards a more sustainable future. **Nucleic Acids Res** 2016; 44(D1): D279–D285. <https://doi.org/10.1093/nar/gkv1344>. PMID: 26673716; PMCID: PMC4702930.
 32. Müller DJ, Wu N, Palczewski K. Vertebrate membrane proteins: structure, function, and insights from biophysical approaches. **Pharmacol Rev** 2008; 60(1): 43–78. <https://doi.org/10.1124/pr.107.07111>. Epub 2008 Mar 5. PMID: 18321962; PMCID: PMC2561275.
 33. Chinchar VG. Replication of viruses. **Encyclopedia of Virology** 1999: 1471–1478. <https://doi.org/10.1006/rwvi.1999.0245>. PMCID: PMC7149704.
 34. Urra H, Dufey E, Lisbona F, Rojas-Rivera D, Hetz C. When ER stress reaches a dead end. **Biochim Biophys Acta** 2013; 1833(12): 3507–3517. <https://doi.org/10.1016/j.bbamcr.2013.07.024>. PMID: 23988738.
 35. Bu Z, Callaway DJ. Proteins move! Protein dynamics and long-range allostery in cell signaling. **Adv Protein Chem Struct Biol** 2011; 83: 163–221. <https://doi.org/10.1016/B978-0-12-381262-9.00005-7>. PMID: 21570668.
 36. Uversky VN. Intrinsically disordered proteins from A to Z. **Int J Biochem Cell Biol** 2011; 43(8): 1090–1103. <https://doi.org/10.1016/j.biocel.2011.04.001>. PMID: 21501695.
 37. Garcés AP, Watowich SJ. Intrinsic flexibility of West Nile virus protease in solution characterized using small-angle X-ray scattering. **Biochemistry** 2013; 52(39): 6856–6865. <https://doi.org/10.1021/bi400782w>. PMID: 24015950.
 38. Enany S. Structural and functional analysis of hypothetical and conserved proteins of *Clostridium tetani*. **J Infect Public Health** 2014; 7(4): 296–307. <https://doi.org/10.1016/j.jiph.2014.02.002>.
 39. Kaur A, Pati PK, Pati AM, Nagpal AK. Physico-chemical characterization and topological analysis of pathogenesis-related proteins from *Arabidopsis thaliana* and *Oryza sativa* using in-silico approaches. **PLoS One** 2020; 15(9): e0239836.
 40. Fernández-Sanlés A, Ríos-Marco P, Romero-López C, Berzal-Herranz A. Functional information stored in the conserved structural RNA domains of flavivirus genomes. **Front Microbiol** 2017; 8. <https://doi.org/10.3389/FMICB.2017.00546>. 546–546.
 41. Rangel MV, Catanzaro N, Thannickal SA, Crotty KA, Noval MG, Johnson KEE, Ghedin E, Lazear HM, Stapleford KA. Structurally conserved domains between flavivirus and alphavirus fusion glycoproteins contribute to replication and infectious-virion production. **J Virol** 2022; 96(2): e0177421. <https://doi.org/10.1128/JVI.01774-21>. PMID: 34757841; PMCID: PMC8791257.
 42. Koo QY, Khan AM, Jung KO, Ramdas S, Miotto O, Tan TW, Brusci V, Salmon J, August JT. Conservation and variability of West Nile virus proteins. **PLoS One** 2009; 4(4): e5352. <https://doi.org/10.1371/journal.pone.0005352>. PMID: 19401763; PMCID: PMC2670515.
 43. Crill WD, Trainor NB, Chang GJ. A detailed mutagenesis study of flavivirus cross-reactive epitopes using West Nile virus-like particles. **J Gen Virol** 2007; 88(4): 1169–1174. <https://doi.org/10.1099/vir.0.82640-0>. PMID: 17374760.
 44. Sharma S, Varani G. NMR structure of Dengue West Nile viruses stem-loop B: a key cis-acting element for flavivirus replication. **Biochem Biophys Res Commun** 2020; 531(4): 522–527. <https://doi.org/10.1016/j.bbrc.2020.07.115>. PMID: 32807496; PMCID: PMC7494574.
 45. Praveen M, Morales-Bayuelo A. Drug designing against VP4, VP7 and NSP4 of rotavirus proteins – in silico studies. **Moroc J Chem** 2023; 11(3): 729–741. <https://doi.org/10.48317/IMIST.PRSM/morjchem-v11i3.40088>.
 46. Sahu LK, Singh K. Cross-variant proof predictive vaccine design based on SARS-CoV-2 spike protein using immunoinformatics approach. **Beni Suf Univ J Basic Appl Sci** 2023; 12(1): 5. <https://doi.org/10.1186/s43088-023-00341-4>. PMID: 36644779; PMCID: PMC9831375.

How to cite this article: Praveen M. Characterizing the West Nile Virus's polyprotein from nucleotide sequence to protein structure – Computational tools. *J Taibah Univ Med Sc* 2024;19(2):338–350.