IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 7 Issue 1, January 2020
ISSN (Online) 2348 – 7968 | Impact Factor (2019) – 6.248
www.ijiset.com

Homology Modelling and Molecular Characterization of E1 structural Glycoprotein involved in Chikungunya

Someshwar M.Moholkar*, Yogesh N. Joshi and Shrutika B.Chintakindi

Dept. of PG Studies and Research in Bioinformatics, Walchand centre for Biotechnology, Solapur, and Maharashtra, India.

E-mail: *sammoholkar95@gmail.com**, *kotashrutika31@gmail.com*

Tel.: +919595469747

Abstract

Chikungunya virus (CHIKV) is a mosquito-transmitted alphavirus that causes high fever, rash, and recurrent arthritis in humans. It has efficiently adapted to Aedes albopictus, which also inhabits temperate regions, including Europe and the United States of America. In the past, CHIKV has mainly affected developing countries, but has recently caused large outbreaks in the Caribbean and Latin America. No treatment or licensed CHIKV vaccine exists. Aedes albopictus is the second-largest transmitter of CHIKV. Other symptoms may occur, including headache, fatigue, digestive complaints, and conjuncvities. E1 glycoprotein mediation of viral membrane fusion during CHIKV infection is a crucial step in the release of viral genome into the host cytoplasm for replication. How the E1 structure determines membrane fusion and whether other CHIKV structural proteins participate in E1 fusion activity remain largely unexplored.

In the present study, we used different *In Silico* tools and techniques which include retrieval of E1 structural glycoprotein sequence from UniProtKB database and the sequence analysis was performed by using Protparam Tool which concluded that the protein was unstable and basic in nature. The secondary structure was predicted by using SOPMA tool which indicated that the percentage of coils was higher than the percentage of alpha helix and extended strand. Then the 3D structure of E1 structural glycoprotein was predicted by using SWISS MODEL server and the model was validated by using PROCHECK analysis After validation of the model, the validation score was 91.2% indicating that the model was of good quality and the predicted 3D structure was deposited in protein model database (PMDB).

Keywords: Chikungunya, E1 structural glycoprotein, UniProtKB and Homology Modelling.

I. INTRODUCTION

Chikungunya is a severe and debilitating disease. Currently, Brazil is experiencing an epidemic caused by three arboviruses, which has changed the way health professionals have diagnosed and treated infected patients. The virus was first described in 1952 during a febrile illness outbreak in Makonde, a province in southern Tanzania [1]. The word chikungunya comes from the Bantu language of the Makonde ethnic group from Tanzania and Mozambique and refers to the curved

position of the patient due to debilitating joint pain [2]. Since its description, in 1952, CHIKV has caused millions of human infections in Africa, the Indian Ocean islands, Asia, Europe, and the Americas[3]. Human CHIKV infection is characterised by an intense joint pain of abrupt onset, high fever, and rash. The infection is self-limited and acute symptoms usually resolve within one–two weeks. Chikungunya is a positive-sense single-stranded RNA virus that is approximately 12 kb in length. The genome has two open reading frames (ORFs): the 5´ORF, translated from genomic RNA, encodes the nsP1, nsP2, nsP3, and nsP4 non-structural proteins, and the 3´ORF, translated from subgenomic RNA, encodes a polyprotein that is processed into the structural proteins [capsid (C), envelope (E1 and E2), and two peptides (E3 and 6K)] [4]. CHIKV is transmitted to humans by mosquitoes of the genus *Aedes* spp, particularly *Aedes aegypti*, which is one of the most efficient mosquito vectors for arboviruses. This efficiency is mainly because this genus is highly anthropophilic and lives in close proximity to humans [5]. *Aedes albopictus* is the second-largest transmitter of CHIKV. A mutation associated with an amino acid substitution in the enveloped glycoprotein (E1–A226V) allowed the virus to better adapt to the vector, thus increasing its ability to transmit and disseminate the virus. This finding was observed in the strain of Chikv that circulated during an outbreak in the Indian Ocean islands, referred to as the Indian Ocean lineage[6]. Other species of mosquitoes, from different parts of the world, have the ability to transmit CHIKV, including *Eretmapodites chrysogaster*, *Culex annulirostris*, *Mansonia uniformis*, *Anopheles stephensi*, and *Opifex fuscus*[5]. Transmission through these vectors is related to their geographical distribution and the types of transmission cycles, whether wild or urban. In the present study, we used different In-Silico tools and techniques for characterization, homology modeling and active site prediction of E1 structural glycoprotein protein. [7] The first step includes retrieval of E1 structural glycoprotein sequence from UniProt KB database. The physiochemical properties were analysed by using ProtParam tool and the secondary structure was predicted by using SOPMA secondary structure prediction tool. Later the 3D structure was predicted by using SWISS-MODEL server and the model was validated by using PROCHECK method.

II. MATERIAL AND METHODS

1. Retrieval of Sequence

The protein sequence of E1 structural glycoprotein was retrieved from UniProtKB protein database and saved in FASTA file format. UniProtKB is a protein sequence database which is freely accessible to the public and it contains the amino acid sequences of proteins [9].

2. Physicochemical Analysis

The physicochemical properties of E1 structural glycoprotein was analyzed by ProtParam analysis tool. The ProtParam tool calculates parameters such as amino acid composition, molecular weight theoretical pI, instability index, aliphatic index and grand average of hydropathicity (GRAVY) [10].

3. Secondary Structure Prediction

The secondary structure was predicted by SOPMA (Self- Optimized Prediction Method with Alignment) method. It was employed for calculating the secondary structural of E1 structural glycoprotein. The SOPMA method correctly predicts the secondary structure α-helix, β-sheet and coil.

4. Homology Modeling and Model Validation

The E1 structural glycoprotein sequence was used for comparative homology modeling using SWISS MODEL server. SWISS-MODEL is a fully automated protein structure homology modeling server to make the protein models accessible to all biotechnologist [11]. After modeling, to check the quality and validation of the model was carried out by PROCHECK Ramchandran plot method using PDBsum server [12].

5. Identification of Domain function and analysis

Identification of domain and its biological function of E1 structural glycoprotein was determined from Pfam domain identification database. The Pfam database is a large collection of protein families, each represented by multiple sequence alignments and hidden Markov models (HMMs).

III. RESULTS AND DISCUSSION

1. Sequence Retrieval

The protein sequence of E1 structural glycoprotein from Chikungunya virus (CHIKV) was retrieved from UniProt KB database and the sequence was saved in FASTA file format in notepad. The protein name, organism name, UniProt KB ID and sequence length were shown in table 1.

Table 1. Retrieval of sequence

Protein name	Organism name	UniprotKB ID	Sequence length
E1 structural glycoprotein	Chikungunya virus (CHIKV)	A0A1E1JNN3	171

2. Physicochemical analysis

The physicochemical properties was analysed by using ProtParam tool and the results were enlisted in table 2. As per the table, the E1 structural glycoprotein is unstable and basic in nature. The total number of positively charged residues (Arg+Lys) was higher than the total number of negatively charged residues (Asp+Glu).

Table 2. Physicochemical parameters of E1 structural glycoprotein

Parameters	Values
Total No. of Amino acids	171
Molecular weight	18181.46 Da
Pi	5.77
Uniprot ID	A0A1E1JNN3
The number of positive amino acids (Arg+Lys)	13
The number of negative amino acids (Asp+Glu)	16
Extinction coefficient	11460
Aliphatic index	73.10

3. Secondary Structure Prediction

The secondary structure was predicted by using SOPMA method. The secondary structure elements like alpha helix, beta sheets, extended strand and random coils were enlisted in Table 3. From the table, the percentage of coils in E1 structural glycoprotein was higher than the percentage of alpha helix and extended strand.

Table 3. Secondary Structure Prediction

Secondary Structure Elements	Number of residues	Percentage
Alpha helix	28	16.37%
Extended Strand	45	26.32%
Beta turn	12	7.02%
Random coil	86	50.29%

4. Homology Modeling and Model Validation

The 3D structure of E1 structural glycoprotein was predicted by the SWISS Model server. The sequence template (Pdb id: 2xfc) E1 Envelope Glycoprotein From Semliki Forest Virus with its 97.08, was selected as template for prediction of homology modelling. The quality and validation of the model was carried out by Ramachandran plot analysis using PDBsum server. Ramachandran plot analysis showed that the percentage of favoured region is 91.2% which was higher than the percentage of additional allowed region it's conclude that the predicted model was reliable and good quality. Further model was visualized in RasMol visualization software package as shown in figure 4.

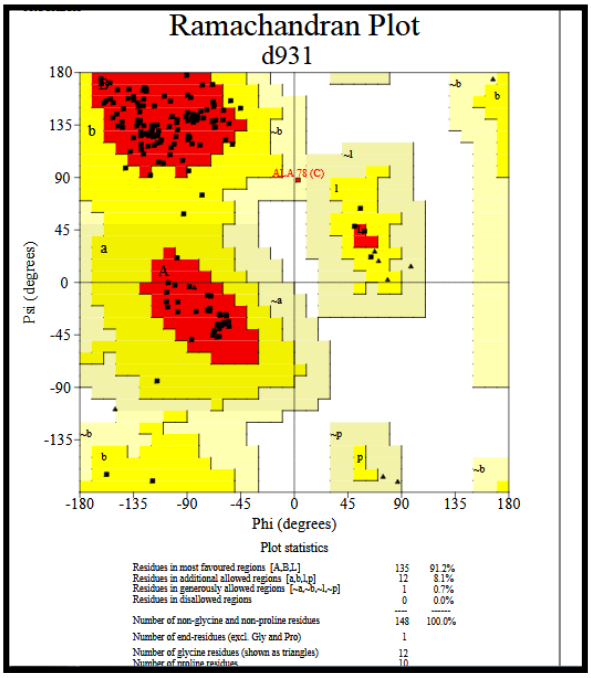

Fig.3 Protein model validation by PROCHECK

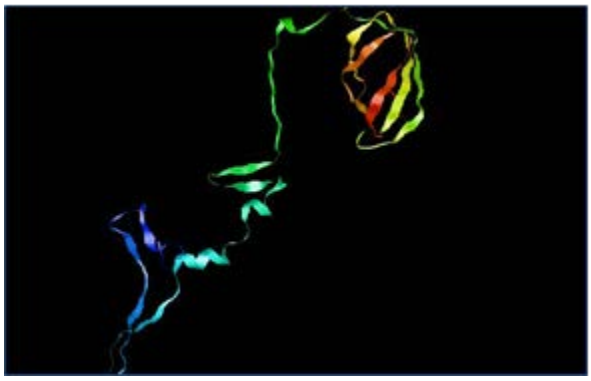

Fig.4 Visulatization of Predicted 3D structure by RasMol

5. Identification of Domain function and analysis

The Protein domain and protein family analysis of E1 structural glycoprotein by Pfam database. The domain of E1 strctural glycoprotein and its sequence length was shown in Fig. 5 which shows it belongs Alphaviruses family. It is like all other group IV viruses, have a positive sense, single-stranded RNA genome. There are thirty alphaviruses able to infect various vertebrates such as humans, rodents, fish, birds, and larger mammals such as horses as well as invertebrates. Transmission between species and individuals occurs mainly via mosquitoes, making the alphaviruses a member of the collection of arboviruses.

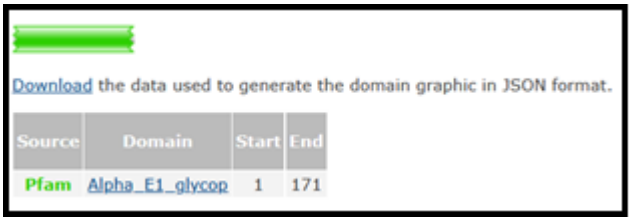

Fig.5 Domain identification with its sequence length

IV. CONCLUSIONS

The present preliminary investigation mainly leads to understand the basic primary, secondary structure and tertiary structure of *28T* E1 structural glycoprotein from Chikungunya virus using various *In-silico* tools and techniques. The E1 structural glycoprotein protein was retrieved from UniprotKB database and it is having length of 171. The primary protein sequence analysis carried out using protparam tool and it retrieval that having a Molecular weight - 18184.46 Da, pI -5.77,The number of positive amino acids (Arg+Lys) -13, The number of negative amino acids (Asp+Glu)-16 respectively., Aliphatic

index-73.10, GRAVY: -0.126. Secondary structure of Chikv protein was predicted by SOMPA having alpha helix-31, beta turns 11, extended strands 42; random coils 87 so protein was highly stable. The Modelling of 3D structure of E1 structural glycoprotein protein was build by Swiss model 2xfc. Structure was validated using PROCHECK having model quality was 91.2%. This present study put molecular insight into the further studies to find the structural and functional properties of this E1 structural glycoprotein to find or design the novel antiviral drugs of Chikungunya virus by Structure
based drug design approach.

V. REFERENCE

[1] Robinson MC. An epidemic of virus disease in Southern Province, Tanganyika Territory, in 1952–53. I. Clinical features. Trans R Soc Trop Med Hyg. 1955; 49 (1): 28-32.
[2] Thiberville S-D, Moyen N, Dupuis-Maguiraga L, Nougairede A, Gould EA, Roques P, et al. Chikungunya fever: epidemiology, clinical syndrome, pathogenesis and therapy. Antiviral Res. 2013; 99(3): 345-70.
[3] Rodrigues FN, Lourenço J, Cerqueira EM, Lima MM, Pybus O, Alcantara LC. Epidemiology of chikungunya virus in Bahia, Brazil, 2014–2015. PLoS Curr. 2016; 1(8):
[4] Schwartz O, Albert ML. Biology and pathogenesis of chikungunya virus. Nat Rev Microbiol. 2010; 8: 491-500.
[5] Coffey L, Failloux A, Weaver C. Chikungunya virus–vector interactions. Viruses. 2014; 6(11): 4628-63.
[6] Tsetsarkin KA, Vanlandingham DL, McGee CE, Higgs S. A single mutation in chikungunya virus affects vector specificity and epidemic potential. PLoS Pathog. 2007; 3: e201
[7] Nunes MRT, Faria NR, Vasconcelos JM, Golding N, Kraemer MUG, Oliveira LF, et al. Emergence and potential for spread of chikungunya virus in Brazil. BMC Med. 2015; 13:102.
[8]Lopes N, Nozawa C. Características geraise epidemiologia dos arbovírus emergentes no Brasil. Rev Pan-Amaz Saude. 2014; 5(3): 55-64.
[9] Magrane M. And the UniProt consortium. UniProt Knowledgebase: a hub of integrated protein data Database, 2011, bar009.
[10] Gasteiger E, Hooglan C, Gattiker A, Duvaud S, Wilkins MR, Appel RD et al. The Proteomics Protocols Handbook, Human Press, 2005, 571-607.
[11] Arnold K, Bordoli L, Kopp J, Schwede T. The SWISSMODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics 2006; 22:195-201.
[12] Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK - a program to check the stereochemical quality of protein structures. J App. Cryst. 1993; 26:283-291
[13] Petersen LR, Powers AM. Chikungunya: epidemiology. F1000Res. 2016; 5(F1000 Faculty Rev): 82.
[14] Schuffenecker I, Iteman I, Michault A, Murri S, Frangeul L, Vaney MC, et al. Genomemicroevolution of chikungunya viruses causing the Indian Ocean outbreak. PLoS Med.2006; 3(7): e263.
[15] Powers AM, and Logue CH. Changing patterns of chikungunya virus: re-emergence of a zoonotic arbovirus. J Gen Virol. 2007; 88(Pt 9): 236377.
[16] Nunes MRT, Faria NR, Vasconcelos JM, Golding N, Kraemer MUG, Oliveira LF, et al.Emergence and potential for spread of chikungunya virus in Brazil. BMC Med. 2015; 13:102.

[17]Carey DE (July 1971). Chikungunya and dengue: a case of mistaken identity. J Hist MedAllied Sci. 26 (3): 243–62. doi:10.1093/jhmas/XXVI.3.243.
[18] The Pfam protein families database in 2019: S. El-Gebali, J. Mistry, A. Bateman, S.R. Eddy, A. Luciani, S.C. Potter, M. Qureshi, L.J. Richardson, G.A. Salazar, A. Smart, E.L.L. Sonnhammer, L. Hirsh, L. Paladin, D. Piovesan, S.C.E. Tosatto, R.D. Finn Nucleic Acids Research (2019) doi: 10.1093/nar/gky995