

# Bioinformatics analysis and characteristics of envelop glycoprotein E epitopes of dengue virus

Hua Zhong<sup>1</sup>, Wei Zhao<sup>1</sup>, Liang Peng<sup>1</sup>, Shan-Feng Li<sup>1</sup>, Hong Cao<sup>1</sup>

<sup>1</sup>Department of Microbiology, School of Public Health and Tropical Medicine, Southern Medical University, Guangzhou, Guangdong 510515, China. Correspondence should be addressed to Hong Cao (gzhaocao@fimmu.com), Tel.: +86 20 61648723.

Received Sep. 4<sup>th</sup>, 2008; revised Dec. 22<sup>nd</sup>, 2008; accepted Jan. 8<sup>th</sup>, 2009

## ABSTRACT

The major envelope glycoprotein E of dengue (DEN) virus plays a central role in the biology of flaviviruses. It is capable of inducing a protective immune response *in vivo* and responsible for the viral binding to the cellular receptor. The crystal structures of glycoprotein E ectodomains have already been determined. However, it is still unclear where the well-defined B-cell epitopes for glycoprotein E which induce the neutralizing antibodies locates. Thus, in order to characterize the role of glycoprotein E in the pathogenesis of dengue virus infection, we first used network servers (<http://bio.dfci.harvard.edu/Tools/> & <http://www.imtech.res.in>) to predict and analyze the well defined B-cell and T-cell epitopes of the glycoprotein of the DEN-1 HAWAII strain. Then based on the highly conserved envelop glycoprotein amino acids, the hydrophilicity, antigenicity, accessibility and flexibility of envelop glycoprotein E were further predicted by using Biotic softwares (DNASTAR) and network servers (<http://bio.dfci.harvard.edu/Tools/>), the secondary structure was putatively obtained. In our study, the sequence at 281-295 amino acid (aa) for dengue virus type 1 HAWAII strain and the sequence at 345-359, 383-397 for dengue virus type 2 NGC strain were predicted as the more prevalent epitopes by using multiple parameters and different analysis softwares, respectively. Two epitopes of DEN-2 and one of DEN-1 locate on the domain III and domain II of the protein E, respectively. Subsequently, further studies will be carried out to examine the antigenicity and protection of the synthetic peptides with higher scores in the average antigen index (AI) and better hydrophilic properties determined by our data.

**Keywords:** Dengue Virus, Glycoprotein E, Epitope, Bioinformatics

## 1. INTRODUCTION

Dengue virus, a flavivirus belonging to the *flaviviridae*

family, is a mosquito-borne human pathogen that causes dengue and dengue hemorrhagic fever which is currently one of the serious public health threats throughout the tropical and subtropical regions of the world [1]. Four serotypes of DEN virus have been identified (DEN-1, 2, 3 and 4), and each of these serotypes can infect humans and cause disease. This virus shares many characteristics with other flaviviruses, having a single-stranded RNA genome surrounded by an icosahedral scaffold and covered by a lipid envelope. The complete virion is 50 nm in diameter and contains an 11-kb plus-sensed RNA genome that is composed of seven nonstructural (NS) protein genes and three structural protein genes, core (C, 100 amino acids), membrane (M, 75 amino acids), and envelope (E, 495 amino acids) [2,3]. The order of proteins encoded is 5'-CprM(M)-E-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5-3'[4]. The 495-amino-acid (aa) envelop (E) glycoprotein, one of the three structural proteins, is the principal component of the external surface of the virion [5], and it is responsible for a wide range of biological activities, including binding to host cell receptors, fusion to and entry into host cells, therefore, this protein directly affects host range, cellular tropism, and, in part, the virulence of the virus [2,5]. Furthermore, the E protein also stimulates host immunity by inducing protective and neutralizing antibodies [6]. It is a main target and important antigen for vaccine development, and many attempts have been made to elucidate the structure-function relationships of the dengue virus glycoprotein E. The crystal structures of protein E ectodomains have already been determined. However, the location of well-defined B-cell and T-cell epitopes for glycoprotein E is largely unknown. Mapping of the B-cell and T-cell epitopes should be important for immunoinfectomic studies of dengue virus infection. Random peptide display has been applied in antigenic epitope determination. However, a combination of computational methods (e.g., bioinformatics) and experimental approaches of conventional biology should be a holistic way to determine the rigorous B-cell and T-cell epitopes. Thus, in order to characterize the role of glycoprotein E in the pathogenesis of dengue virus infection, we used bioinformatics and molecular approaches to predict and analyze its B-cell and T-cell antigen epitopes. Parameters

such as hydrophilicity, flexibility, accessibility, turns, exposed surface, polarity and antigenic propensity of polypeptide chains have been correlated with the location of continuous epitopes in a few well-characterized proteins. Net servers and the software DNA star are applied in our study.

## 2. MATERIALS AND METHODS

### 2.1. Antigenic Peptide Prediction

The online web server (<http://bio.dfci.harvard.edu/Tools/>) give us a pathway to predict sequences of peptides within a protein that are likely to be antigenic by eliciting an antibody response. Antigenic peptides are determined using the method of Kolaskar and Tongaonkar [7]. Predictions are based on a table that reflects the occurrence of amino acid residues in experimentally known segmental epitopes. We enter the amino acids of dengue virus type 1 HAWAII strain as well as dengue virus type 2 NGC (New Guinea C) strain, both of whom are standard strains, then operate the applet.

### 2.2. Hydrophilicity Estimation

The website locating at (<http://us.expasy.org/tools/protscale.html>) can give us a hydrophilicity prediction of the envelop glycoprotein E, it is based on the method of Hopp & woods.

### 2.3. Secondary Structure Presumption

Logging in the same web server mentioned in the third step, the  $\beta$ -turn and coil of the E protein can be obtained, using the algorithm [8,9] of Levvit as well as Deleage & Roux [10,11].

### 2.4. Surface Accessibility and Average Flexibility Assumption

The Protean procedure of The DNASTAR software can supply us with the E protein's Surface accessibility and flexibility using methods of Emini [12] and Karplus-Schulz [13].

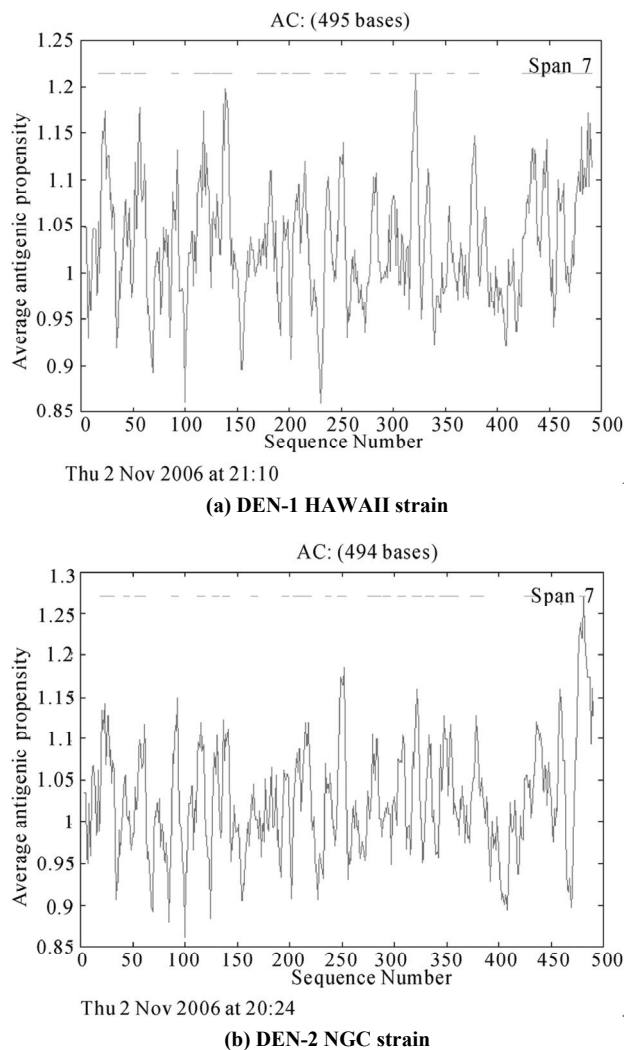
### 2.5. Tertiary Structure Prediction

After the process of secondary structure prediction, we can use the method of JPRED to link in the PDB ID, since the sequence of the protein E has 39 hits to the PDB database with E values of less than 0.0001. Thus, we make an entry in the PDB in the ID "IP58" to have a look at the structure by using the viewer Software named RASWIN32b2a.

## 3. RESULTS

### 3.1. Antigenic Peptide Prediction of the Glycoprotein E

The prediction results for antigenic peptides of the glycoprotein E for DEN-1 and DEN-2 are shown in **Figure 1**.



**Figure 1.** Peptides predicted as antigenic epitope sites of the protein E.

**Table 1.** Peptides predicted as B-cell epitope sites of the protein E.

No	Start position	Sequence	End position
1	17	GATWVDVLEHGSCVT	32
2	39	PTLDIELLKT	48
3	51	TNPAVLRKLCIE	62
4	87	DANFVCRR	94
5	109	GKGLITCAKFKCVTK	124
6	126	EGKIVQYENLKYSVIVTHT	145
7	169	PTSEIQLTDYGALTLDCSP	187
8	193	FNEMVLL	199
9	204	KSWLVHKQWFLDLPLPW	220
10	234	EDLLVTFKT	242
11	246	KKQEVAVLG	254
12	278	IFAGHLKCRLL	287
13	296	GMSYVMCTG	304
14	316	QHGTVLVQVK	325
15	329	TDAPCKIPF	337
16	352	ITANPIVT	359
17	373	FGESYIVVGA	382
18	424	SIGGVFTSVGKLVHQIFGTAYGVLFSG	450
19	458	GIGILLTW	465
20	472	SASLSMTCIAVGMVTLYLGV	491

The B-cell (**Table 1**) and T-cell epitopes (**Table 2**) of the glycoprotein of DEN-1 HAWAII standard strain were predicted by the means of the position of amino acids with the online server respectively.

### 3.2. Hydrophilicity Prediction of the Glycoprotein E

The prediction results for hydrophilicity of the protein E of DEN-1 and DEN-2 are diagramed in **Figure 2**.

**Table 2.** Peptides predicted as T-cell epitope sites of the protein E.

HLA Sites	Peptides Position		
HLA-A2	206-215	117-126	483-492
HLA-A11	299-308	238-247	50-59
HLA-A24	298-307	439-446	325-334
HLA-B51	216-225	420-446	206-215
HLA-B60	48-57	313-322	256-265
HLA-B62	414-423	291-300	124-133

### 3.3. Secondary Structure Prediction

The prediction of protein E's secondary structure is shown in **Figure 3**.

### 3.4. Surface Probability of the Glycoprotein E

The surface probability assumption of the protein E of DEN-1 and DEN-2 strains is diagramed in **Figure 4**.

### 3.5. Flexibility Presumption of the Glycoprotein E

The Flexibility presumption of the glycoprotein E of DEN-1 and DEN-2 strains is shown in **Figure 5**.

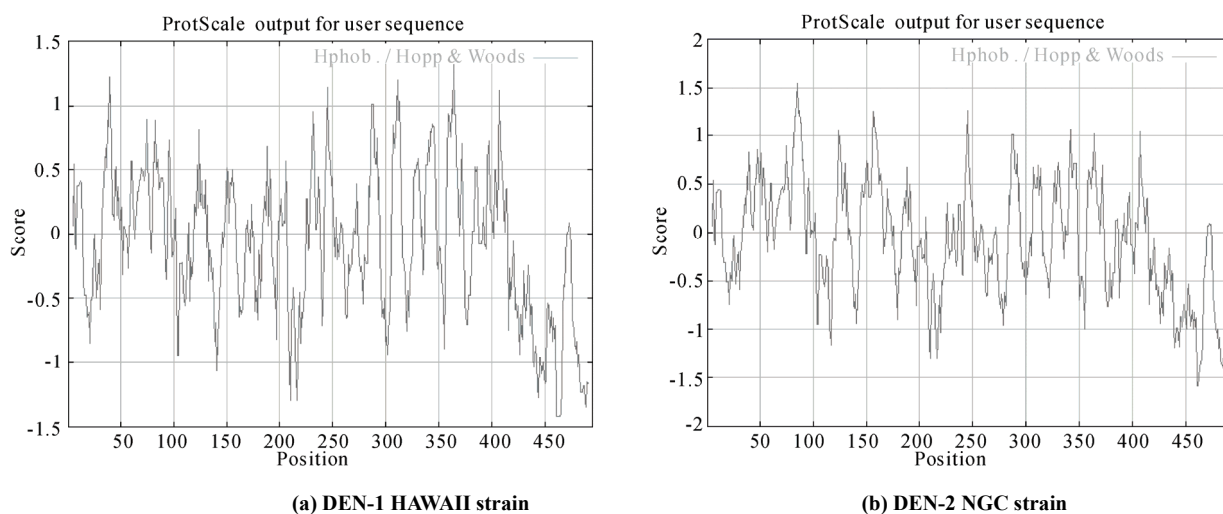
### 3.6. Putative Tertiary Structure of the Glycoprotein E

The putative tertiary structure of the glycoprotein E is shown in **Figure 6** (protein data bank). The ID is IP58.

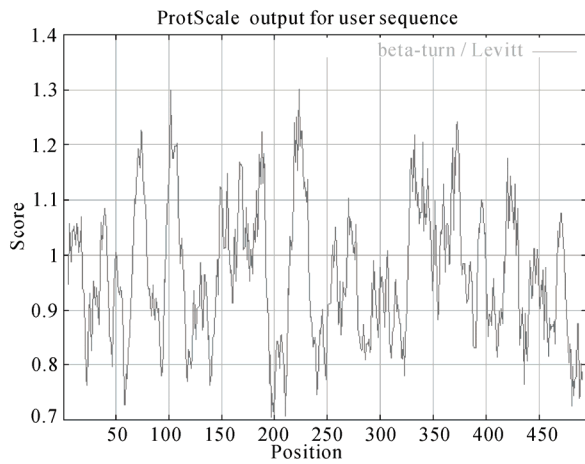
## 4. DISCUSSION

It is well known that although urgently needed for dengue virus infection, specific drugs for treatment and effective vaccination for prevention are currently unavailable. This is the main reason why we have focused on the so-called Antibody-dependent enhancement (ADE). Studies suggest that during a secondary infection with a different serotype, the presence of cross-reactive, non-neutralizing antibodies enhances the efficiency with which dengue virus infects susceptible cells. A molecular understanding of the events that lead to antibody neutralization, enhancement, or escape will be critical to the improvement of vaccines. It is therefore important to determine which surface features on the dengue virion are responsible for inducing protective or enhancing immune response in the different serotypes. Thus, the structural and functional organizations of the dengue virus proteins are of central interests for the understanding of the biology of dengue virus and the mechanisms of virus-cell interactions.

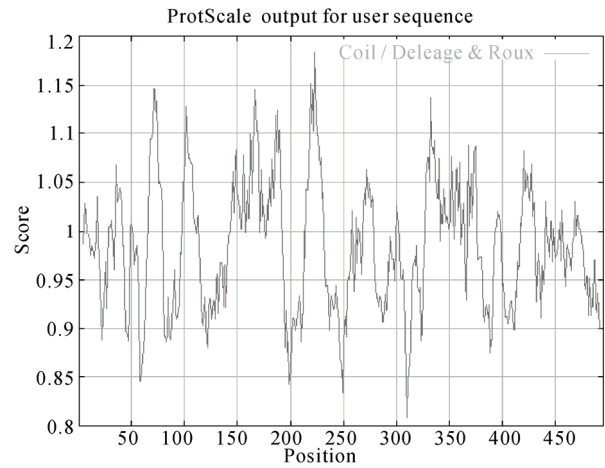
The dengue E ectodomain consists of structurally distinct domains: I, II and III [14]. The domain III appears to play an important part in host cell receptor binding for viral entry and in inducing protective immunity. The rigorous B-cell and T-cell epitopes were not identified yet. In our study, we focused on the characterizing the B-cell and T-cell epitopes of dengue virus envelop E glycoprotein by deploying the bioinformatics approaches, the sequence at 281-295 amino acid (aa) for dengue virus type 1 HAWAII strain and the sequence at 345-359, 383-397 for dengue virus type 2 NGC strain were predicted as the more prevalent epitopes by using multiple parameters and different analysis softwares, respectively. The sequences selected not only have higher scores in the average antigen index (AI), which could predict the antigen epitope of envelop glycoprotein E, but also showed better hydrophilic properties. Two epitopes of DEN-2 and one of DEN-1 locate on



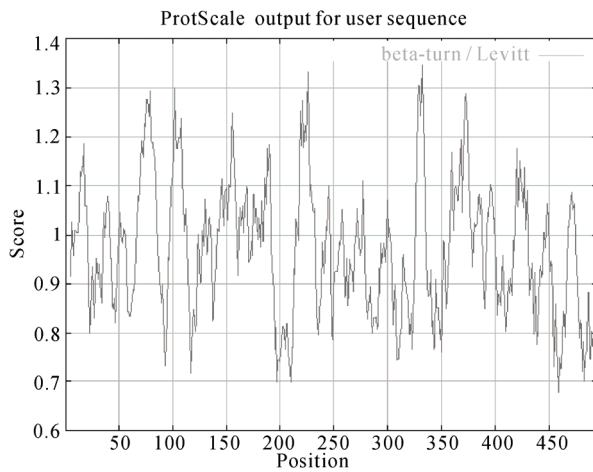
**Figure 2.** The hydrophilicity of the protein E.



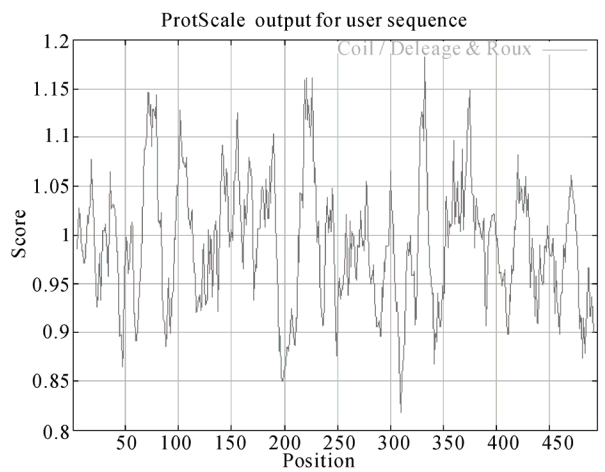
(a1)  $\beta$ -turn the coiled region of DEN-1 HAWAII strain



(a2)  $\beta$ -turn the coiled region of DEN-1 HAWAII strain

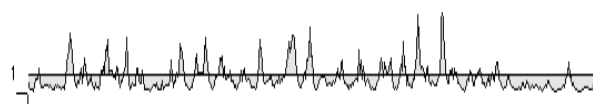


(b1)  $\beta$ -turn the coiled region of DEN-2 NGC strain

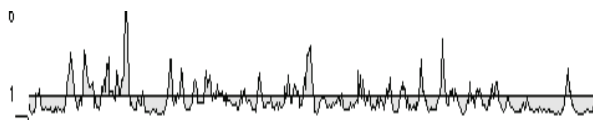


(b2)  $\beta$ -turn the coiled region of DEN-2 NGC strain

**Figure 3.** The prediction of the secondary structure of the protein E.



(a) DEN-1 HAWAII strain



(b) DEN-2 NGC strain

**Figure 4.** Surface probability of the glycoprotein E.

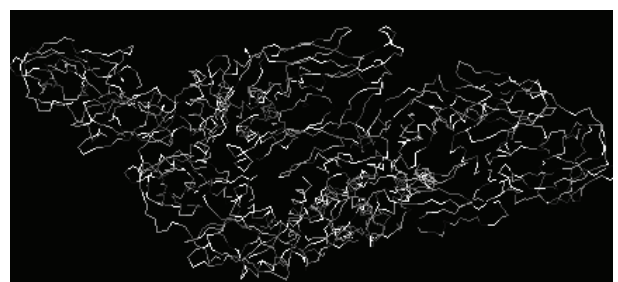


(a) DEN-1 HAWAII strain



(b) DEN-2 NGC strain

**Figure 5.** The flexibility presumption of the glycoprotein E.



**Figure 6.** Putative tertiary structure of the glycoprotein E.

the domain III and domain II of the protein E, respectively. The domain III has been hypothesized to contain multiple type- and subtype-specific epitopes eliciting only virus-neutralizing monoclonal antibodies while the domain II is involved in virus-mediated membrane fusion, and contains many cross-reactive epitopes eliciting neutralizing and non-neutralizing monoclonal antibodies. The predicted epitopes can be used for the devel-

opment of vaccine and the dissection of the ADE effect. The further experimental studies will be performed to determine the immunogenicity and protection effect of peptides with higher scores in the average antigen index (AI) and better hydrophilic properties, and to identify vaccine candidates.

## ACKNOWLEDGEMENTS

This work was supported by Guangzhou Key Technology R&D Program (No.2008Z1-E401 to H. Cao).

## REFERENCES

- [1] T. Monath. (1994) Dengue: the risk to developed and developing countries. *Proc Natl Acad Sci USA* 91, 2395–2400.
- [2] R. J. Kuhn, W. Zhang, M. G. Rossmann. (2002) Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. *Cell* 108, 717–725.
- [3] E. A. Henchal, J. R. Putnak. (1990) The dengue viruses. *Clin Microbiol Rev* 3, 376–396.
- [4] C. M. Rice, B. N. Fields, D. M. Knipe, P. M. Howley. (1996) *Flaviviridae: the viruses and their replication*. *Virology* 3, 931–959.
- [5] J. T. Roehrig. Immunochemistry of the dengue viruses, 199–219 *In: D. J. Gubler and G. Kuno (eds), Dengue and dengue hemorrhagic fever*. CAB International, New York, N. Y. 1997.
- [6] G. J. Chang. Molecular biology of dengue viruses, 175–198 *In: D. J. Gubler, G. Kuno (eds.), Dengue and Dengue Hemorrhagic Fever*. CAB International, London, 1997.
- [7] A. S. Kolaskar, P. C. Tongaonkar. (1990) A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett* 276, 172–174.
- [8] M. Levitt, J. Greer. (1977) Automatic identification of secondary structure in globular proteins. *J Mol Biol* 114, 181–239.
- [9] J. Cheng. (2008) A multi-template combination algorithm for protein comparative modeling. *BMC Structural Biology* 8, 18–36.
- [10] G. Deleage, B. Roux. (1987) An algorithm for protein secondary structure prediction based on class prediction. *Protein Eng* 1, 289–294.
- [11] J. Martin, J. F. Gibrat, F. Rodolphe. (2006) Analysis of an optimal hidden Markov model for secondary structure prediction. *BMC Structural Biology* 6, 25–45.
- [12] E. A. Emini, J. V. Hughes, D. S. Peflow. (1985) Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol* 55, 836–839.
- [13] P. A. Karplus, G. Schultz. (1985) Prediction of chain flexibility in proteins. *Naturwissenschaften* 72, 212–213.
- [14] Y. Modis, S. Ogata, D. Clements. (2003) A ligand-binding pockets in the dengue virus envelop glycoprotein. *Science* 100, 6986–6991.