The Bioinformatics Analysis of Hepatitis C Virus E2 Protein

Tailin Guo Shuhui Wang Yihong Wu

School of Bioengineering, Southwest Jiaotong University, Chengdu 610031, China

Abstract

By the methods of bioinformatics, the sequences of HCV E2 protein from all the genotypes are studied on the variation and the B-cell epitopes. We find that even in the HVR1 and HVR2 regions, there are also some conservative sites, such as T2, G6, G23, and Q26, all of which belong to polar R-base amino acids without charge, and can lost proton to conform hydrogen bond. Meanwhile, we find the conservative amino acids locate on the surface of E2 protein, make up of β-turn and B-cell epitopes. This will contribute to prompt the development of HCV vaccine.

Keywords: Hepatitis C Virus, Bioinformatics, E2 protein, Epitopes, Conservative sites

1. Instruction

Hepatitis C virus (HCV) is the major cause of acute and chronic infectious hepatitis, affecting more than 170 million people worldwide [1]. Over 80% of acutely infected individuals progress to a chronic carrier state that can lead to liver cirrhosis and hepatocellular carcinoma [2, 3]. In acute HCV infection, an early HCV-specific cellular immune response is associated with viral clearance and recovery [4, 5], whereas in chronically infected individuals, cellular immune responses are generally low and unable to eliminate the virus [6]. No prophylactic HCV vaccine is currently available and the only accepted therapy thus far, interferon-a, is successful in only 20% of the cases [7]. Increased efforts are therefore needed in the development of an effective vaccine against HCV.

E2 protein is one of the glycoprotein, which was reported to play and important role in the HCV attachment and entry into the target cells. Antibodies targeted against the N-terminal 27-amino-acid (aa) region of E2 (aa 384 to 411), which is the most ariable region (known as hypervariable region 1 [HVR-1]) of the HCV polyprotein, inhibit the binding of glycoprotein E2 to cells and block HCV infectivity in vitro and in vivo (8, 9). Therefore it is one of the most valuable candicate proteins for the development of HCV vaccine. However, on account of its high mutation rate, the traditional methods failed. With the development of bioinformatics, we have some tools and methods to study the rules in the variation of E2 protein based on that as an important structural protein, there must be some conservative region to maintain its function and structure, and by the analysis of great amount of sequence of E2 protein, we maybe find some rules in the variation and then predict the future sequence and structure to construct new vaccine for the preventing of HCV infection.

2. aa sequences compare of E2

In the web of NCBI, we searched 1000 sequences of E2 protein and download to lenovo server. The average number of amino acids is 367. Besides, we got 18 sequences from the most representational genotypes which was annotated in Hadeberger conference in 2004 [10]. The sequences were also used in the other parts of the paper.

By the soft of Clustal W in GCG package, more than one thousand aa sequences were compared with Lenovo server in our lab as shown in the protocols to find the mutational /conservative regions and sites, and we find that in E2 protein, there are two hypervariable regions, HVR1(aa 1 to 27) and HVR2 (aa 95 to 103) [13]-[15], which are consistent with the ones shown in the papers. Besides, we also find hyper-conservative sites of aa, such as aa 2, aa 24, aa 102, et al.

3. Evaluation of entropy power

Entropy power stand the uncertainty of some site, and the more the number is,the more the aa mutates.by the function of entropy plot in BIOEDIT soft package, all the entropy power of aas in E2 protein were evaluated.

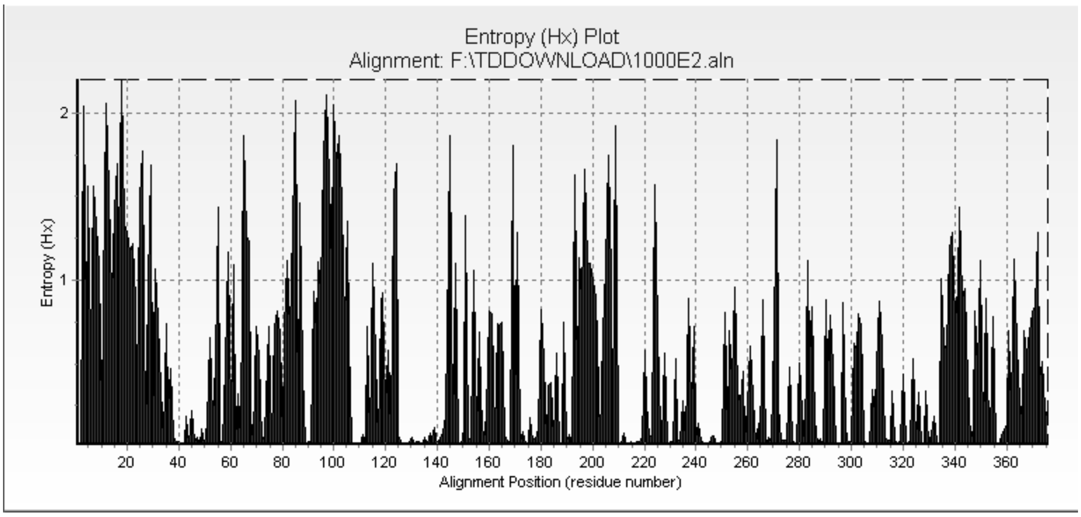

Fig. 1: the entropy power result of E2 protein by BIOEDIT.

As shown in Fig 1, there are several peaks,which contains HVR1 (aa 1 to 20) and HVR2(aa 95 to 100). So entropy power can be used to screen the variable regions.

4. Structural prediction

4.1 Secondary structural prediction

All the E2 proteins sequences from six genotypes was uploaded to the internet server, and secondary structures was predicted automatically using SOPMA library, and the address is (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_hnn.html). There are not clear difference in the six genotypes, and structure of E2 protein belong genotype 1 are shown in Fig 2.

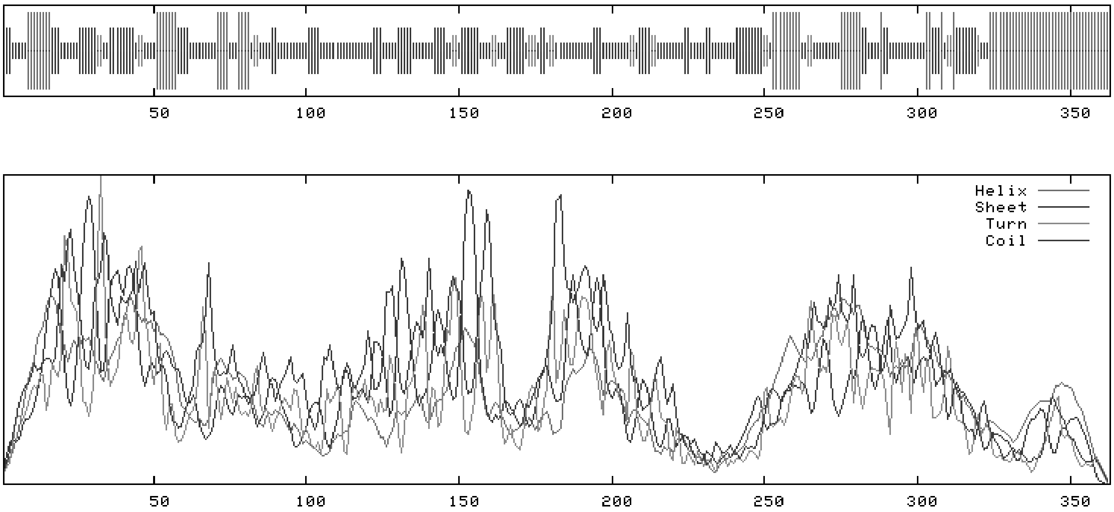

Fig. 2 E2 protein illustration of secondary structure

In the results, we can see thatα-helix only lie in the regions of aa 9 to 15, aa 50 to 60, aa260 to 267, aa280 to 287 and aa 300 to 384, and few ß-turns spread all the sequence. Just because there are not great differences in different genotypes, we think it were the conservative sites that maintain the special structure of E2 proteins. while the structure of ß-folding spreads in the sequence equably,

4.2 Tretyi structural prediction

In the server of SWISS-PROT, we find one template which matches with E2 protein, and the E2 sequences were uploaded to the server to draw the tretyi structural illustrations of E2 protein with the softs of swiss-pdbviewer and rasmol, and to find out the functional motif.

Meanwhile, SCRATCH servers (http://www.igb.uci.edu/), developed by California university, was used to protein structure prediction by the methods of PSI-BLAST and neural network, in which under the limitation of geometry, the secondary structure, analysissitus characters and detailed residues information are used.

In the tretyi structurue of E2 protein, there are 2 α-helix, and several β-turns which are accompanied with the results of secondary structure, but there is notβ-folding. the possible reason is that the sequence is not long enough and there are lack of similar structure information in the PDB database.

However, all the methods are not good enough to predict the tretyi structural of E2. Using the tretyi structural template of HCV, and the part of E2 was selected to be analyzed by the soft of Jmol on the structure and conservative.

From the prediction results of secondary and trytyi structure, we find that the conservative sites of aa 2,6,24,35-50,104-110,131 are one part of β-ture and even β-hairpin.

A. β-turn beside aa 6 ;B. hairpin beside aa 44 to 50 ;C.β-turn and hairpin beside aa 131

5. Prediction of B-cell epitopes

The characters of hydrophilicity, plasticity, Exterior accessibility, antigenicity and secondary structure used in the prediction of E2 protein epitopes[11,12] on account that just one of them is not

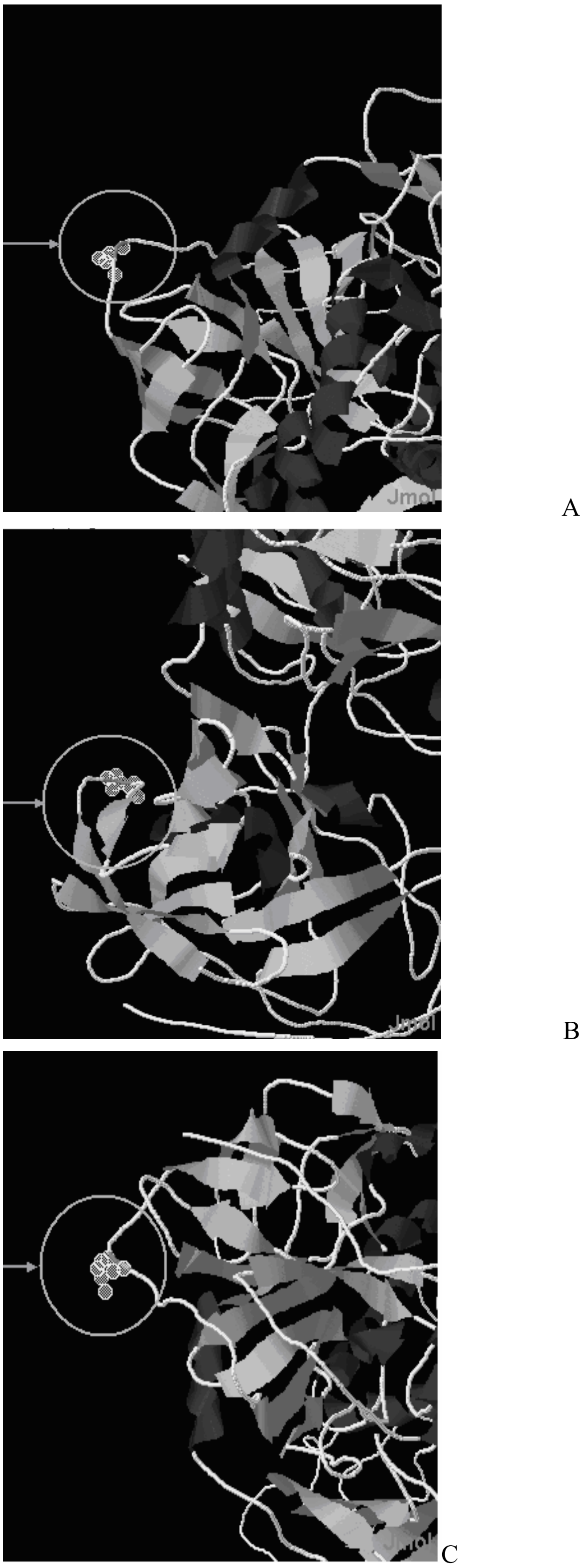

Fig. 3: The tretyi structure of E2 protein by Jmol.

enough to get the good results,and that the epitopes locate on the surface of protein,facilitate the combination with antibody,and is flexibility generally.

The program of Protean was selected in Dnastar soft package.Firstly,the epitopes just based on one character,including Hydrophilicity(Hopp Woods and Kyte Doolittle), Antigenicity (Jameson-wolf), Flexibility (Karplus-Schulz),Surface Probability(Emin) and Secondary structure (Garnier-Robson) were predicted respectively. the results of genotypes are shown in Table 1.

• Prediced parameters	• Epitopes
• hydrophilicity	• 1-12, 22-31, 33-39, 44-52, 60-84, 87-92, 94-109, 136-150, 158-164, 174-177, 201-217, 219-222, 225-227, 229-238, 251-279,
• plasticity	• 5-14, 21-27, 32-36, 41-42, 47-52, 64-72, 78-81, 85-88, 93-101, 107-110, 117-118, 127-130, 134-143, 149-152, 157-164, 173-178, 190-195, 199-201, 205-219, 264-269, 272-281, 285-287, 303-304, 323-325
• Exterior accessibility	• 9-12, 22-27, 33-35, 49-50, 60-65, 72, 78-79, 97-101, 106-109, 135-144, 149-151, 157-162, 176, 205-210, 227-233, 264-278,
• antigenicity	• 1-13, 21-27, 32-36, 45-53, 61-81, 92-103, 107-112, 134-143, 147-153, 157-166, 173-177, 192-194, 200-220, 261-282

Table 1: predicted parameters of genotype 1.

The same methods were used for the other genetyps, and finally all the results were combined with β-turn to give the most possible epitopes.As shown in Table 2.

• genotype	• Epitopes
• 1	• 9-11, 22-24, 33-35, 49-50, 64-72, 98-101, 107-109, 136-143, 149-150, 157-164, 174-177, 208-209, 264-269, 273-278
• 2	• 14, 33-35, 49, 65-72, 93-94, 97-102, 110-111, 139-140, 144-145, 160-162, 164-166, 176-179, 209-213, 268-271, 277-282
• 3	• 22-24, 33-35, 49-50, 64-65, 97-102, 110-111, 138-144, 150-152, 162-166, 178, 191-198, 211-215, 225, 269-274, 277-281
• 4	• 22-24, 33-35, 49, 64-65, 79-83, 96-101, 158-163, 176-178, 190-194, 206-209, 264-265, 274-276
• 5	• 22-25, 33-35, 49-50, 64-65, 69-74, 96-102, 109-110, 138-144, 147-151, 162-164, 176-178, 191-197, 206-210, 267-269, 273-277
• 6	• 11, 21-23, 32-34, 48, 63-64, 82-84, 92-100, 137-143, 162-163, 174-177, 192-197, 212-213, 268-271, 277-280

Table2: predicted epitopes of different genotypes.

Although there are great mutation in the sequence of E2 protein, we find some similar antigen epitopes, as shown in Table 2, among which, the most visible ones are aa 11to 24, aa 28 to 35, aa 48 to 50, aa 64 to 74, aa 96 to 110, aa 137 to 148, aa 176 to 178, aa191 to 197, aa208 to 225, aa 264 to 277. We believe that one of the sequence maybe conformed the conservative part of one epitopes.

6. Discussion

By the compare of more than one thousond of E2 protein sequences, we found that :there are two hyper variable regions which named HVR1(aa 1 to 127) and HVR2 (aa 95 to 103). However, the E2 protein still stable to a certain extent, 64.8% of the bases is conserved, the mutation rate in the unhyper variable region is only 25.5%, and even in the hyper variable regions. There are some conservative sites, such as aa T2 ,G6,G23,Q26. Besides, it is still conservative both in glycosylation sites(131) and phosphorylation sites (aa66,aa129,aa131...). Meanwhile, we also found that there is few Cys, Trp, Asp and Met, while great number of Gly, Thr and Ser, which are all polar amino acids. We speculated that all of these related with the B-cell epitopes.theoretically, E2 protein locates on the surface of viral particle, and there must be some epitopes of B-cells. So, in order to escape the human being immuno-system, there are great mutation sites. However, it play great roles in the entrance of viral into cells by anchoring at CD81 [16]. So the structure must be maintained and there must be some rules in the amino acids mutation. Our results prove it and some details should be demonstrated by further study.

In the prediction of protein structure, we found that there are β-turn and β-hairpin which are stable structure beside conservative sites, such as No.2,6,24,35-50,104-110,131. So we speculated that it is these sites that maintain the stable structures of E2 protein, and the conservative sites were the basic bearing sites for the construction of E2 protein tretyi structure. Further study should focus on that if amino acids of the conservative sites were placed, then how to change about the tretyi structure?

On the other hands, we studied the change of entropy power at different part of E2 protein using soft of BIOEDIT. The abscissa is the sequence of E2 protein, coordinate is the value of entropy power, and more great the entropy power value is, more higher the variation frequency is. So this is good methods in the studying of E2 mutation in different regions. As the results shown in figures, there are great accordance with the former results.

Finally, based on the study of mutation, structural prediction and entropy power about E2 protein, we predicted the B-cell epitopes of E2 protein. Since Hopp and Woods developed a method to predict B-cell epitopes with hydrophilicity parameters in 1980,s, more and more methods have been provided, and among which, the most acceptable ones are Hydrophilicity method, Accessibility method, Antigenicity method, Flexibility method and secondary structure method.

On the base of prediction of secondary structure, Hydrophilicity method, Accessibility method, Antigenicity method and Flexibility method were selected to predict the epitopes. After four sets of results were integrated, we find most possible epitopes of E2 protein as shown in the following: aa 11 to 24, aa 28 to35, aa48 to 50, aa 64 to 74, aa 96 to110, aa137to148, aa176to178, aa191to197, 208to225, 264to27. All the results demonstrated that though E2 sequence is the most variable in HCV genome, we find some conservative sites in different genotypes and predict the possible B-cell epitopes, which will prevent the development HCV vaccine.

Envelope glycoprotein of HCV is always the hot sites which attracts many researchers attention, not only there is neutralize antigen epitopes which can induce antibodies to prevent HCV infection in HVR1 and HVR2, but also the sequence is hyper variable which caused antigen drift to make virus escape from host immunosystem and replicate sostenuto. In the past papers, we find little on the conservative sites in the hyper variable regions, and few systematic statistics, while in our study, we searched a great deal of sequences and analyzed them systematacially. We find that even in the HVR1 and HVR2 regions, there are also some conservative sites, such as T2,G6,G23,and Q26, all of which belong to polar R-base amino acids without charge, and can lost proton to conform hydrogen bond. Meanwhile, we find the conservative amino acids locate on the surface of E2 protein, make up of β-turn and B-cell epitopes. We believe that all of these finding will prompt the development of HCV vaccine and more research should focus on the structure of E2 and the interaction with CD81 and antibody with the method of molecular dynamics and molecular modeling.

References

[1] Q. L. Choo, G. Kuo, A.J.Weiner, Overby, Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome, *Science* ,pp244: 359 – 362,1988.

[2] H. J. Alter, L. B. Seeff, Recovery, persistence, and sequelae in hepatitis C virus infection: a perspective on long-term outcome. Semin. *Liver Diseases* ,20: 17 – 35,2000.

[3] S. Cooper, et al., Analysis of a successful immune response against hepatitis C virus, *Immunity*,10: 439 – 449,1999.

[4] N. H.Gruner, et al, Association of hepatitis C virus-specific CD8+ T cells with viral clearance in acute hepatitis C, *Journal Of Infective. Diseases* ,181: 1528 – 1536 ,2000.

[5] J.G. McHutchison, et al, Interferon alpha-2b alone or in combination with ribavirin as initial treatment for chronic hepatitis C: Hepatitis Interventional Therapy Group, *New England Journal of Medicine* , 339: 1485 – 1492 ,1998.

[6] Q.L. Choo, et al, Genetic organization and diversity of the hepatitis C virus, *Proceedings of National Academic Science*. USA , 88: 2451 – 2455,1991.

[7] A. Grakoui, C. Wychowski, , C.Lin, , S.M. Feinstone and C. M. Rice, Expression and identification of hepatitis C virus polyprotein cleavage products. *Journal of Virology*. 67: 1385 – 1395, 1993.

[8] F.Habersetzer, A. Fournillier and J. Dubuisson, Characterization of human monoclonal antibodies specific to the hepatitis C virus glycoprotein E2 with in vitro binding neutralization properties, *Virology*, 249:32–41,1998.

[9] D.Rosa, S. Campagnoli and C. Moretto, A quantitative test to estimate neutralizing antibodies to the hepatitis C virus:

cytofluorimetric assessment of envelope glycoprotein 2 binding to target cells *Proceedings of National Academic Science*. USA ,93:1759–1763,1996.

[10] P. Simmonds, J. Bukh, C. Combet, et al, Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes, *Hepatology*, 42:962-973, 2005.

[11] M.P. Manns and E.G. Rambusch, *Journal of Hepatology*, 31:39-42, 1999.

[12] A. Fournillier, C. Wychowski, D. Boucreux, et al. Induction of hepa-titis C virus E1 envelope protein-specific immune response can be enhanced by mutation of N-glycosylation sites, *Journal of Virology*, 75:12088-12097, 2001.

[13] F. Penin, C. Combet, G. Germanidis, et al. Conservation of the conformation and positive charges of hepatitis C virus E2 envelope glycoprotein hypervariable region 1 points to a role in cell attachment, *Journal of Virology*, 75: 5703-5710, 2001.

[14] M. Hijikata, N. Kato, Y. Ootsuyama, et al. Hypervariable regions in the putative glycoprotein of hepatitis C virus. *Biochemical and Biophysical Research Communications*, 175:220-228, 1991.

[15] Y. Suzuki, T.Gojobori, Positively selected amino acid sites in the entire coding region of hepatitis C virus subtype 1b, *Gene*, 276: 83-87, 2001.

[16] M. Lechmann, T.J. Liang, Vaccine development for hepatitis C, *Seminars in Liver Disease*, 20(2):211-226,2000.