Construction and Cloning of Structural Protein Gene of Hepatitis C Virus Genotype 1b In Escherichia coli and Its Expression In Chinese Hamster Ovary Cells

Conference Paper · August 2014

CITATIONS
0

READS
584

3 authors:

Hidayat Setiadji
Biofarma
2 PUBLICATIONS **0** CITATIONS
SEE PROFILE

Nurainy Neni
Biofarma
37 PUBLICATIONS **179** CITATIONS
SEE PROFILE

Debbie Retnoningrum
Bandung Institute of Technology
59 PUBLICATIONS **514** CITATIONS
SEE PROFILE

Construction and Cloning of Structural Protein Gene of Hepatitis C Virus Genotype 1b In *Escherichia coli* and Its Expression In Chinese Hamster Ovary Cells

Debbie S Retnoningrum[1]*, Hidayat Setiadji[1,2]*§, Neni Nurainy[2]*

[1]Shcool of Pharmacy, Bandung Institute of Technology, Bandung, West Java, Indonesia

[2]Department of Research, BFM, Bandung, West Java, Indonesia

*These authors contributed equally to this work

[§]Corresponding author

Email addresses:

DSR: debbieretnoningrum@yahoo.com

HS: ajisakti2003@students.itb.ac.id and ajisekolah@gmail.com

NN: ainy_nur@yahoo.com

Abstract

Background

Hepatitis C Virus (HCV) is a member of Hepacivirus genus in Flaviviridae family which genome consisting of positive single stranded RNA molecule. Polyprotein precursor encoded by this RNA would be cleaved by co-translational and post-translational process to generate structural (core and envelope, C, E1, E2) and non-structural proteins. This study was aimed to obtain recombinant HCV structural protein as a vaccine candidate that has $CD4^+$ T cells and $CD8^+$ T cells immune response, and could be manufactured in production scale. In order to obtain HCV structural protein as a vaccine component, the nucleotide sequence of C, E1, E2 (CE1E2) proteins was constructed based on consensus of amino acid sequences of CE1E2 proteins of genotype 1b which was derived from multiple sequence alignment of amino acid sequences that were retrieved from several web sites.

Results

We analysed over 550 publically available amino acid sequences of CE1E2 protein of genotype 1b to determine consensus sequence of CE1E2 protein. The consensus sequence was selected automatically by multiple sequence alignment using both Clustal X software and manual alignment. We found that consensus amino acid sequence has a GenBank accession number ABV46113.1. We also identified a consensus sequence belonging to big group of amino acid sequence as showed by phylogenetic tree. Through analysis of signal peptide by Phobius software we found that consensus sequence didn't have any signal peptide sequence. We also identified that consensus sequence has linear epitope of MHC I and MHC II and conformational epitope of B cell.

Of Conclusion

Consensus sequence of CE1E2 protein could represent amino acid sequences of CE1E2 protein of genotype 1b. CE1E2 protein could be expressed as intracellular protein in Chinese Hamster Ovary (CHO) cells and predicted to have $CD4^+$ T cells, $CD8^+$ T cells and B cells immune response.

Keywords

HCV, Core, E1, E2, pVAX1, CHO.

Background

Hepatitis C Virus (HCV) infection prevalence average ranting 2%-3% of world population. Chronic HCV infection could cause cirhosis, hepatic fibrosis and develop hepatic failure and cancer [1]. Naturally HCV infection could not give sufficience immune response. Current therapeutic drug could only treat viral infection on few patients with a serious side effect. Therefore, vaccine development is needed to prevent and overcome an infection wirh cross neutralization, envelope protein antibody and $CD4^+$ T cells and $CD8^+$ T cells immune response. An effective HCV vaccine require a multi component epitope that could rise variable protective immunity. Strategic based on polyepitope of HCV-*like particles* consisting of CE1E2

protein (structural protein) could meet an effective vaccine criteria. Structural protein is located on N terminal of HCV polyprotein, whether nun-structural protein is located on C terminal. Structural polypeptide is cleaved by sigma peptida peptidase of host endoplasmic reticulum (ER) [2]. Major function of core protein is to form capsid to protect HCV RNA genome [3,4,5]. Two envelope glycoproteins, E1 and E2, are important component for viral entry and fusion. These protein have several function such as membran attachment, ER localization and heterodimer assembly. E1 and E2 proteins are highly glycocilated, consisting of five to 11 glycocilation sites. E2 has a hypervariable region with amino acid sequence that could be different up to 80% among HCV genotype and among subtype of the same genotype [6]. Hiper Variable Region 1 (HVR1) has 27 amino acid and neutralization epitope [7]. Inspite of HVR1 variable sequence, physicochemical properties of residue on each position and whole HVR1 conformation are highly conserved among all known HCV genotype, indicating an important role on viral life cycle [8]. HVR1 is a basic region with positive charge residue which interacted with negative charge molecule on cell surface. This interaction play roles in host cell recognition and attachment [9]. HCV envelope glycoproteins do not require a cleaving by host cell protease, during transport on secretion tract [10]. C proteins multimerize and form a capsid to encapsidate HCV RNA on cytoplasmic surface of ER. Established capsid in cytoplasm interact with E1. When expressed on variable in vitro system, including mamalian cell culture, bacteria, insect, or yeast, C proteins could establish nucleocapsid-like particles. Many frame-shift during C region translation express F protein, which antibody of this protein circulate in chronic liver infected patients [11]. Transmission electron microscopy observation give images of capsid without envelope in cytoplasm of hepatic cell, whereas particle with an envelope has captured in cisternae of ER [12,13].

HCV develop several mechanism to survive in human host cell and to evasive from cellular and adaptive immunity. HCV has an ability to inhibit natural killer cell activity [14,15], interferon type 1 induction [16] and to produce mutant which could escape from *cytotoxic T lymphocytes* [17]. Viral structural protein which is expressed in convenient heterolog system could form capsid or envelope that could self assembly to be Virus Like Particle (VLP) without another viral component [18]. In this study, consensus of amino acid sequence of CE1E2 protein is used to design a nucleotide sequence of CE1E2 protein that could be expressed in CHO cells. The expressed CE1E2 proteins are expected to give cellular and humoral immune response.

Results

Design of nucleotide sequence of encoded gene of CE1E2 protein

At least 100.000 HCV amino acid sequences were analysed to find an amino acid sequence of C, E1 and E2 protein. A number of 4.188 amino acid sequences of CE1E2 protein from several web sites were collected. Multiple sequence alignments of these amino acid sequences from several groups of Clustal X alignment's were joined [19]. A quantity of 594 amino acid sequences of of CE1E2 protein of HCV genotype 1b were separated to give consensus of amino acid sequence by Bioedit software version 7.2.5 [20]. An amino acid sequence with NCBI GenBank accession number's ABV46113.1 (sequence number 55552) was elected as a consensus sequence after manual alignment and sorting each sequence by summed residue

frequencies in selected column from C terminal to N terminal. A position of consensus amino acid sequence among 4.187 amino acid sequences based on amino acid sequence of C protein could be seen in a phylogenetic tree by SeaView software [19,21,22,23,24] (Figure 1).

Figure 1 Phylogenetic tree of amino acid sequence of CE1E2 protein based on C protein. A total of 4.188 amino acid sequences were multiple aligned based on C protein. A consensus of amino acid sequence indicated by number 55552.

Signal peptide analysis using Phobius software give an information that consensus amino acid sequence of CE1E2 protein do not have any signal peptide sequence (Figure 2). Consensus amino acid sequence was analysed for linear epitope MHC I. A linear epitope MHC I, PKPQRKTKRN, was found in C protein region of consensus amino acid sequence. A number of 15 linear epitope MHC II were found in region C, E1 and E2 proteins region (Table 1). Five B cell linear epitope (Table 2) and 29 conformational epitope (Table 3) were also found in consensus amino acid sequence. Conformational epitope position of consensus amino acid sequence is located by using three dimensional of consensus protein (Figure 3). Consensus CE1E2 protein position to natural proteins were compared using E2 protein (PDB ID. 4MWF:C) and C protein (PDB ID. 1XCT:P) (Figure 4).

Figure 2 Signal peptide analysis graphic of consensus amino acid sequence. Graphic is showed five transmembran regions, three cytoplasmic regions and two non-cytoplasmic region.

Table 1 Linear epitope MHC II of consensus amino acid sequence

Epitope	Position	Amino acid sequence
VSTNPKPQRKTKRNT	Protein C	2 – 16
STNPKPQRKTKRNTN	Protein C	3 – 17
IPKARQPEGRAWAQP	Protein C	66 – 80
PKARQPEGRAWAQPG	Protein C	67 – 81
KARQPEGRAWAQPGY	Protein C	68 – 82
LSPRGSRPSWGPTDP	Protein C	99 – 113
SPRGSRPSWGPTDPR	Protein C	100 – 114
PRGSRPSWGPTDPRR	Protein C	101 – 115
*GVDG*ETTVTGGEAGR	Protein E1E2	381 – 395
*VDG*ETTVTGGEAGRT	Protein E1E2	382 – 396
*DG*ETTVTGGEAGRTT	Protein E1E2	383 – 397
TKTCGGPPCNIGGVG	Protein E2	562 – 576
TCGGPPCNIGGVGNN	Protein E2	564 – 578
CGGPPCNIGGVGNNN	Protein E2	565 – 579
TRGERCDLEDRDRSE	Protein E2	649 – 663

Italic letter represent amino acid residue of E1 protein

Table 2 B cell linear epitope of consensus amino acid sequence

No	Start residue	End residue	Amino acid sequence
1.	657	748	EDRDRSELSPLLLSTTEWQILPCSFTTLPALSTGLIHL HQNIVDVQYLYGIGSAVVSFAIKWEYVLLLFLLLAD ARVCACLWMMLLIAQAEA
2.	264	364	DLLVGAATFCSAMYVGDLCGSVFLVSQLFTFSPRRY ETVQDCNCSIYPGHVSGHRMAWDMMMNWSPTAAL VVSQLLRIPQAVVDMVAGAHWGVLAGLAYYS
3.	83	190	PWPLYGNEGMGWAGWLLSPRGSRPSWGPTDPRRRS RNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARALA HGVRVLEDGVNYATGNLPGCSFSIFLLALLSCLTIPT
4.	483	559	QRPYCWHYAPRPCGIVPAAEVCGPVYCFTPSPVVVG TTDRFGVPTYRWGENETDVLLLNNTRPPLGNWFGC TWMNST
5.	476	479	DEHN

Table 3 B cell conformational epitope of consensus amino acid sequence

No	Amino acid residue
1.	Y720, V721, L722, L723, L724
2.	F725, L726, L727, L728, A729, D730, A731, R732, V733, C734, A735, C736, L737
3.	W738, M739, M740, L741, L742, I743, A744, Q745, A746, E747, A748
4.	Y703, L704, Y705, G706, I707, G708, S709, A710, V711, V712, S713, F714, A715, I716, K717
5.	H695, Q696, N697, I698, V699, D700, V701, Q702
6.	W321, D322, M323, M324
7.	T682, T683, L684, P685, A686, L687, S688, T689, G690, L691, I692, H693, L694
8.	R297, R298, Y299, E300, T301, V302, Q303, D304, C305, N306, C307, S308, I309, Y310, P311, G312, H313, V314, S315, G316, H317, R318
9.	Q675, I676, L677, P678, C679, S680, F681
10.	G103, S104, R105, P106, S107
11.	R157, V158, L159, E160, D161, G162, V163, N164, Y165
12.	W108, G109, P110, T111, D112, P113, R114
13.	R116, S117, R118
14.	N119, L120, G121, K122, V123, I124, D125, T126, L127, T128, C129, G130, F131, A132, D133, L134, M135, G136, Y137, I138, P139, L140, V141, G142, A143, P144, L145, G146, G147, A148, A149, R150, A151, L152, A153, H154, G155, V156
15.	L669, S670, T671, T672, E673, W674
16.	M325, N326, W327, S328, P329, T330, A331, A332, L333, V334, V335, S336, Q337, L338, L339, R340, I341, P342, Q343, A344, V345, V346, D347, M348, V349, A350, G351, A352, H353, W354, G355, V356, L357, A358, G359, L360, A361, Y362
17.	Y87, G88, N89, E90, G91, M92, G93, W94, A95, G96, W97, L98, L99, S100, P101, R102
18.	R522, F523, G524, V525, P526, T527, Y528
19.	Y277, V278, G279, D280, L281, C282, G283, S284, V285, F286, L287, V288,

No	Amino acid residue
	S289, Q290, L291, F292, T293, F294, S295, P296
20.	F510, T511, P512, S513, P514, V515, V516, V517, G518, T519, T520, D521
21.	R493, P494, C495, G496, I497, V498, P499, A500, A501, E502, V503, C504, G505, P506, V507, Y508, C509
22.	A166, T167, G168, N169, L170, P171, G172, C173, S174, F175, S176, I177, F178, L179, L180, A181, L182, L183, S184, C185, L186, T187, I188, P189, T190
23.	R659, D660, R661, S662, E663, L664, S665, P666, L667, L668
24.	W488, H489, Y490, A491, P492
25.	D264, L265, L266, V267, G268, A269, A270, T271, F272, C273, S274, A275, M276
26.	W530, G531, E532, N533, E534, T535, D536, V537, L538, L539, L540, N541, N542, T543, R544, P545, P546, L547, G548, N549, W550, F551, G552, C553, T554
27.	W84, P85, L86
28.	M556, N557, S558, T559
29.	Q483, R484, P485

Note : A (Alanine, Ala), R (Arginine, Arg), N (Asparagine, Asn), D (Aspartic acid, Asp, C (Cysteine, Cys), Q (Glutamine, Gln), E (Glutamic acid, Glu), G (Glycine, Gly), H (Histidine, His), I (Isoleusine, Ile), L (Leusine, Leu), K (Lysine, Lys), M (Metionine, Met), F (Phenilalanine, Phe), P (Proline, Pro), S (Serine, Ser), T (Threonine, Thr), W (Tryptophan, Trp), Y (Tyrosin, Tyr), dan V (Valine, Val).

Figure 3 Visualisation of B cell conformational epitope of consensus CE1E2 protein. A total of 29 conformational epitopes are indicated by red color. N-terminal is on midle-left (Met-1) and C-terminal on upper left (Ala-748).

Figure 4 Visualisation of consensus CE1E2 protein position to natural protein. Pictures on left side are superimpose of three dimensional consensus CE1E2 protein (purple) position to natural E2 protein, PDB ID. 4MWF:C (a) and to natural C protein, PDB ID. 1XCT:P (b).

Nucleotide sequence of encoded gene of CE1E2 protein was designed using CHO codon preference from consensus amino acid sequence of CE1E2 protein. Kozak sequence, start and stop codon were added to nucleotide sequence (Figure 5) as it would be expressed in CHO cells using pVAX1 expression vector.

Figure 5 Nucleotide sequence of encoded gene of consensus CE1E2 protein. Consensus amino acid sequence of CE1E2 protein is translated by using CHO codon preference.

Cloning of nucleotide sequence of encoded gene of consensus CE1E2 protein in expression vector

Recombinant vector, pUC-CE1E2 and expression vector, pVAX1© (Figure 6) that have been cloned in *E. coli* were isolated by using QIAprep® Spin Miniprep kit. Recombinant plasmids, pUC-CE1E2 were digested by *Bam*HI and *Pst*I enzyme to give four DNA fragments theoreticaly, sized 2263 bp, 106 bp, 29 bp, and 3236 bp.

Three band were observed on electroforegram, sized 2000 bp, 3000 bp and 5000 bp (Figure 7). DNA insert was confirmed by *Eco*RV enzyme digestion that give two bands, sized 2000 bp and 3000 bp (Figure 8).

Figure 6 Vector map for pUC-CE1E2 and pVAX1©. Recombinant vector, pUC-CE1E2 has encoded gene of kanamycin resistance, universal priming site M13 and *BamH*I and *Pst*I restriction sites. Expression vector, pVAX1© has encoded gene of kanamycin resistance, universal CMV and T7 forward and BGH reverse priming site and *BamH*I and *Pst*I restriction sites.

Figure 7 Electroforegram of pUC-CE1E2 and pVAX1© that have been digested by restriction enzyme. Lane 1, 1 kb marker; lane 2, digested pUC-CE1E2, sized 2000 bp, 3000 bp, and 5000 bp; lane 3, digested pVAX1©, sized 3000 bp.

Figure 8 Vector map for pVAX1©-CE1E2 and electroforegram of pVAX1©-CE1E2 analysis. Lane 1, 1 kb marker; lane 2, amplification of encoded gene of CE1E2 protein from pVAX1©-CE1E2; lane 3, digested pVAX1© by *Eco*RV restriction enzyme.

Expression of encoded gene of consensus CE1E2 protein

Expression of structural protein HCV was characterized by internal celullar staining methods using Fluorescence-Activated Cell Sorting (FACS). A little level expression of CE1E2 protein was observed as much as 6.73% on 24 hours. On 48 hours, level expression was observed as much as 2.04%. Whether on 72 hours there is no expression observed (Figure 9).

Figure 9 Observation of expression level of CE1E2 protein. Percentage of expression level on 24 hours, 6.73%; percentage of expression level on 48 hours, 2.04%; no significant expression level on 72 hours.

Discussion

Design of nucleotide sequence of encoded gene of CE1E2 protein

A multiple sequence alignment suggest that C protein is highly conserved and could be use as a marker for phylogenetic tree. A consensus amino acid sequence is on a big group of C protein that could represent amino acid sequence of HCV CE1E2 protein genotype 1b. As consensus amino acid sequence lack of signal peptide, CE1E2 protein would be expressed intracellular. Consensus amino acid sequence is predicted by epitope analysis to give T cell $CD8^+$ and $CD4^+$ immune response and also B cell immune response. Furthermore, visualisation of conformational epitope and superimpose three dimensional structure of consensus CE1E2 protein strenghtened the prediction. Consensus amino acid sequence translation using CHO codon preference give a nucleotide sequence with codon adaptation index value, 0.72.

Cloning of nucleotide sequence of encoded gene of consensus CE1E2 protein in expression vector

Electroforegram of pUC-CE1E2 and pVAX1© digestion by *Bam*HI and *Pst*I enzyme explain undigested pUC-CE1E2, sized 5746 bp and undigested pVAX1©, sized 2999 bp. Electroforegram of pVAX1©-CE1E2 digestion by *Eco*RV enzyme give two bands

that close to theoritical DNA fragment size. Expression vector pVAX1© is choosen because it's construction fulfill Food and Drug Administrtion, Points to Consider on Plasmid DNA Vaccines for Preventive Infectious Disease Indications.

Expression of encoded gene of consensus CE1E2 protein

Low level expression of encoded gene of consensus CE1E2 protein because complex ratio recombinant vector and Lipofectamine® is not optimal. The incubation time requirement for complex establishment should be optimized with green fluorescent protein.

Conclusions

In conclusion, Multiple sequence alignment using both Clustal X software and manual alignment have choosen consensus amino acid sequence which has a GenBank accession number ABV46113.1. Consensus sequence of CE1E2 protein could represent amino acid sequences of CE1E2 protein of genotype 1b. Consensus amino acid sequence translation using CHO codon preference give a nucleotide sequence with codon adaptation index value, 0.72. CE1E2 protein could be expressed as intracellular protein in Chinese Hamster Ovary (CHO) cells and predicted to have $CD4^+$ T cells, $CD8^+$ T cells and B cells immune response.

Methods

Design of nucleotide sequence of encoded gene of CE1E2 protein

Amino acid sequences of CE1E2 protein from web site www.ncbi.nlm.nih.gov, http://www.uniprot.org/uniprot and http://www.Viprbrc.org/brc are used to do multiple sequence alignment by BioEdit software. Multiple sequence alignment was done by Clustal X software. The Consensus sequence was elected by amino acid sorting according to summed residue frequencies in each column from C terminal to N terminal of amino acid sequences of CE1E2 protein. Consensus amino acid sequence was analyzed for signal peptide by Phobius prediction software and homology analysis by BLAST online software. Nucleotide sequence of consensus amino acid sequence was generated by Optimizer and Spreadsheet CodonOpt software by considerig codon preference for CHO K1 (*Cricetulus griseus*) cell. Codon adaptation index was determined from http://www. genscript.com/cgi-bin/tools/. Consensus amino acid sequence position among another amino acid sequences could be seen by generating phylogenetic tree based on C protein as a marker. Consensus amino acid sequence was further epitope analysed for T cell MHC I and MHC II by IEDB software to see whether the consensus sequence has an epitope that could give T cell $CD4^+$ and $CD8^+$ immune response. Nucleotide sequence of encoded gene of CE1E2 was synthesized by IDT DNA company in pUC recombinant vector.

Cloning of nucleotide sequence of encoded gene of consensus CE1E2 protein in expression vector

Cloning of nucleotide sequence of encoded gene of consensus CE1E2 protein was done by transformation in *E. coli* TOP 10 competent host cells. Heatshock was elected for transformation method. The mixture of recombinant vector and *E. coli* TOP 10 competent host cells was incubated in 37°C for an hour by shaking. Transformants were inoculated in agar plate with 100 µg/mL kanamycin concentration and incubated in 37°C for 14 – 18 hours. One colony of *E. coli* TOP 10

was then inoculated in LB medium with 100 µg/mL kanamycin concentration. Plasmids were isolated by kit QIAprep® Spin Miniprep procedure.
DNA insert was ligated in expression vector, pVAX1© by T4 ligase enzyme. Plasmid concentration was measured by kit Qubit™ dsDNA BR Assay. DNA insert in expression vector was confirmed by restriction analysis using *Eco*RV restriction enzyme. DNA insert was also confirmed by PCR to give DNA fragment, sized 2277 bp.

Expression of encoded gene of consensus latiCE1E2 protein

Recombinant vector, pVAX1©-CE1E2 was complexed with Lipofectamine® LTX and PLUS™ reagent. DNA recombinant complex was transfected in CHO K1 cells. CE1E2 protein expression level was observed by intra cellular staining in 24, 48 and 72 hours. CHO K1 cells have a total quantity not less than 10^6 cells/mL. CHO K1 cells were transfered into F-12 Ham's media in 25 cm^2 culture flask which would be stored in 37°C incubator with 5% CO_2 concentration. CHO K1 cells were passaged after reaching 100% confluent. CE1E2 protein is confirmed by intra cellular staining by indirect staining using anti-C protein mouse IgG1 compared with isotype mouse IgG1 FITC (Figure 10).

Figure 10 Illustration of intra cellular staining of CE1E2 protein in CHO K1 cells. Direct staining of negative control (a) and indirect staining CE1E2 protein (b).

Availability of supporting data

The amino acid sequences of CE1E2 protein available in this article are publically available online as described. Additional files with supporting results are included with this article.

Authors' contributions

NN and DSR conceived the idea for this article. HS design nucleotide sequence of encoded gene of CE1E2 protein based on collected amino acid sequences from several web sites, clone and express it in CHO K1 cells. HS wrote the manuscript. All authors read and approved the final submission.

Acknowledgements

DSR is funded by Bandung Institute of Technology. NN and HS are funded by BFM.

References

1. Nelson PK: **Global Epidemiology of Hepatitis B and Hepatitis C in People Who Inject Drugs: Results of Systematic Reviews**. *Lancet* 2011, **378** (9791):571–583.

2. Plotkin S, Orenstein W, and Offit P: *Vaccines*, Saunders Elsevier, fifth ed., 2008, 47:1187-1196.

3. Chang SC, Yen JH, Kang HY, Jang MH, and Chang MF: **Nuclear Localization Signals in the Core Protein of Hepatitis C Virus**. *Biochem Biophys Res Commun* 1994, **205**:1284-1290.
4. Suzuki R, Matsuura Y, Suzuki T, Ando A, Chiba J, Harada S, Saito I, and Miyamura T: **Nuclear Localization of the Truncated Hepatitis C Virus Core Protein with its Hydrophobic C Terminus Deleted**, *J Gen Virol* 1995, **76**:53-61.
5. Suzuki R, Sakamoto S, Tsutsumi T, Rikimaru A, Tanaka K, Shimoike T, Moriishi K, Iwasaki T, Mizumoto K, and Matsuura Y: **Molecular Determinants for Subcellular Localization of Hepatitis C Virus Core Protein**, *J Virol* 2005, **79**:1271-1281.
6. Weiner AJ, Paliard X, and Selby MJ: **Intrahepatic Genetic Inoculation of Hepatitis C Virus RNA Confers Cross-protective Immunity**, *J Virol* 2001, **75**:7142-7148.
7. Zibert A, Kraas W, Meisel H, Jung G, and Roggendorf M: **Epitope Mapping of Antibodies Directed Against Hypervariable Region 1 in Acute Self Limiting and Chronic Infections Due to Hepatitis C Virus**, *J Virol* 1997, **71**:4123-4127.
8. Penin F, Combet C, Germanidis G, Frainais PO, Deleage G, and Pawlotsky JM: **Conservation of the Conformation and Positive Charges of Hepatitis C Virus E2 Envelope Glycoprotein Hypervariable Region 1 Points to a Role in Cell Attachment**, *J Virol* 2001, **75**:5703-5710.
9. Barth H, Schafer C, Adah MI, Zhang F, Linhardt RJ, Toyoda H, Kinoshita-Toyoda A, Toida T, Van Kuppevelt TH, and Depla E: **Cellular Binding of**

Hepatitis C Virus Envelope Glycoprotein E2 Requires Cell Surface Heparan Sulfate, *J Biol Chem* 2003, **278**:41003-41012.

10. Op DBA, Voisset C, Bartosch B, Ciczora Y, Cocquerel L, Keck Z, Foung S, Cosset FL, and Dubuisson J: **Characterization of functionalhepatitis C virus envelope glycoproteins**. *J Virol* 2004, **78**:2994-3002.
11. Branch AD, Stump DD, and Gutierrez JA: **The Hepatitis C Virus Alternate Reading Frame (ARF) and its Family of Novel Products: The Alternate Reading Frame Protein/ F-protein, the Double-frameshift Protein, Others**, *Semin Liver Dis* 2005, **25**:105-117.
12. Dubuisson J, Hsu HH, and Cheung RC: **Formation and Intracellular Localization of Hepatitis C Virus Envelope Glycoprotein Complexes Expressed by Recombinant Vaccinia and Sindbis Viruses**, *J Virol* 1994, **68**:6147-6160.
13. Dubuisson J: **Folding, Assembly and Subcellular Localization of Hepatitis C Virus Glycoproteins**, *Curr Top Microbiol Immunol* 2000, **242**:135-148.
14. Crotta S, Stilla A, and Wack A: **Inhibition of Natural Killer Cells Through Engagement of CD81 by the Major Hepatitis C Virus Envelope Protein**, *J Exp Med* 2002, **195**:35-41.
15. Tseng CT, and Klimpel G: **Binding of the Hepatitis C Virus Envelope Protein E2 to CD81 Inhibits Natural Killer Cell Functions**, *J Exp Med* 2002, **195**: 43-49.
16. Foy E, Li K, and Sumpter RJ: **Control of Antiviral Defenses Through Hepatitis C Virus Disruption of Retinoic Acid-Inducible Gene-I Signaling**, *Proc Natl Acad Sci USA* 2005, **102**: 2986-2991.

17. Erickson AL, Kimura Y, and Igarashi S: **The Outcome of Hepatitis C Virus Infection is Predicted by Escape Mutations in Epitopes Targeted by Cytotoxic T Lymphocytes**, *Immunity* 2001, **15**: 883-895.
18. Palucha A, Loniewska A, Satheshkumar S, Chachulska AMB, Umashankar M, Milner M, Haenni AL, Savithri HS: **Virus-Like Particles: Models for Assembly Studies and Foreign Epitope Carriers**, *Progress in Nucleic Acid Research and Molecular Biology* 2005, **80**:135-68.
19. Sievers F: **Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega**. *Molecular Systems Biology* 2011, **7**:539.
20. Hall TA: **BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT**. *Nucl. Acids. Symp. Ser*. 1999, **41**:95-98.
21. Edgar RC: **MUSCLE: Multiple Sequence Alignment with High Accu-racy and High Throughput**, *Nucleic Acids Res*. 2004, **32**(5):1792-1797.
22. Castresana J: **Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Molecular Biology and Evolution** 2000, **17**(4):540-552.
23. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, and Gascuel O: **New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0**. *Systematic Biology* 2010, **59**(3):307-321.
24. Felsenstein J: *PHYLIP version 3.5*, 1993.

Figures

See additional file 1

Figure 1 Phylogenetic tree of amino acid sequence of CE1E2 protein based on C protein. A total of 4.188 amino acid sequences were multiple aligned based on C protein. A consensus of amino acid sequence indicated by number 55552.

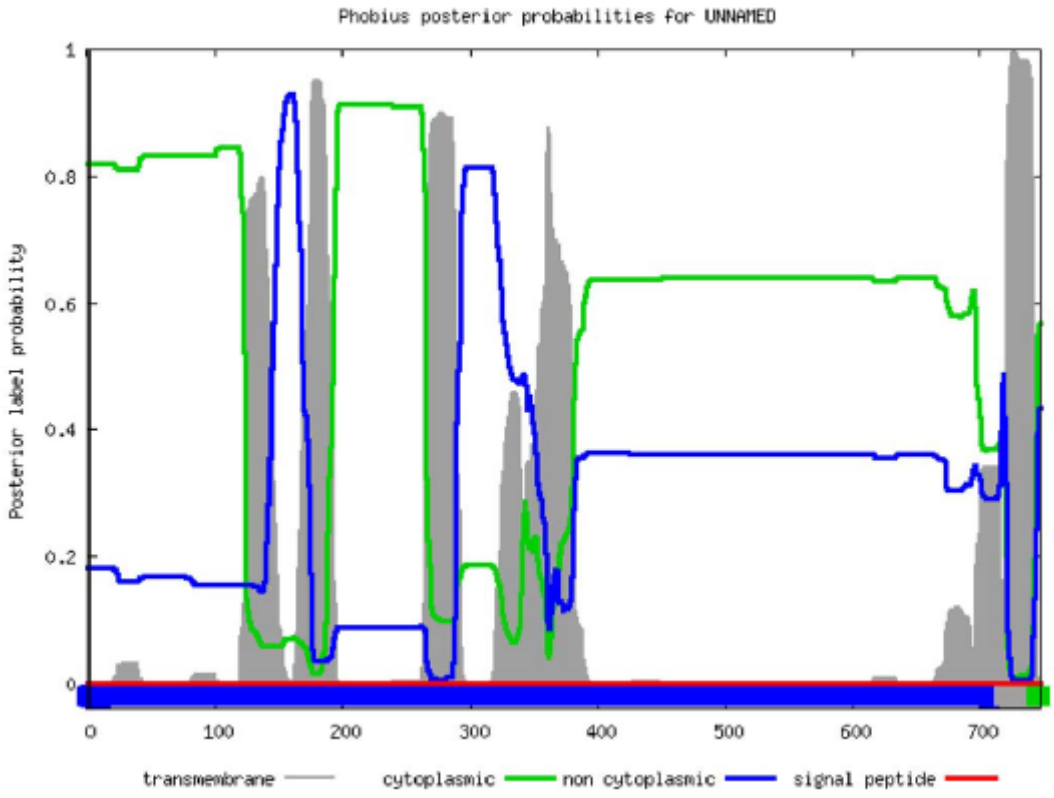


Figure 2 Signal peptide analysis graphic of consensus amino acid sequence. Graphic is showed five transmembran regions, three cytoplasmic regions and two non-cytoplasmic region.

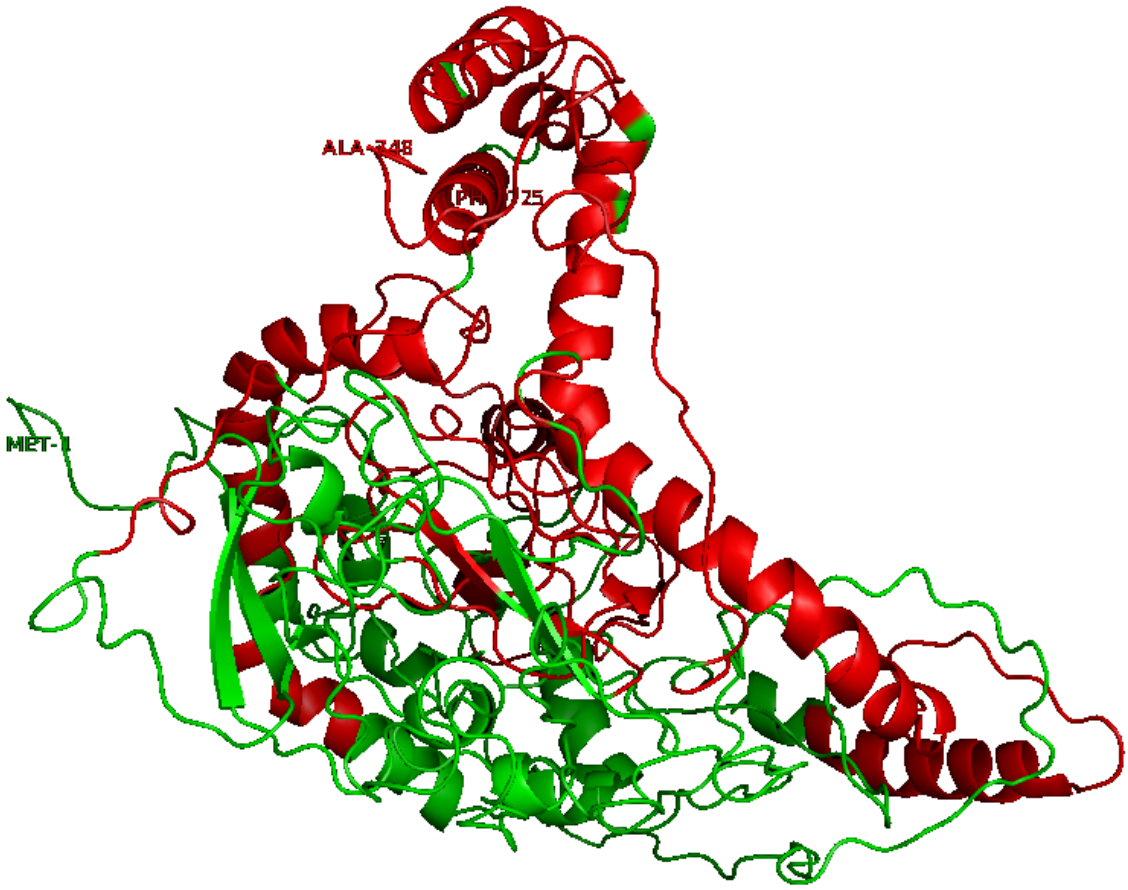
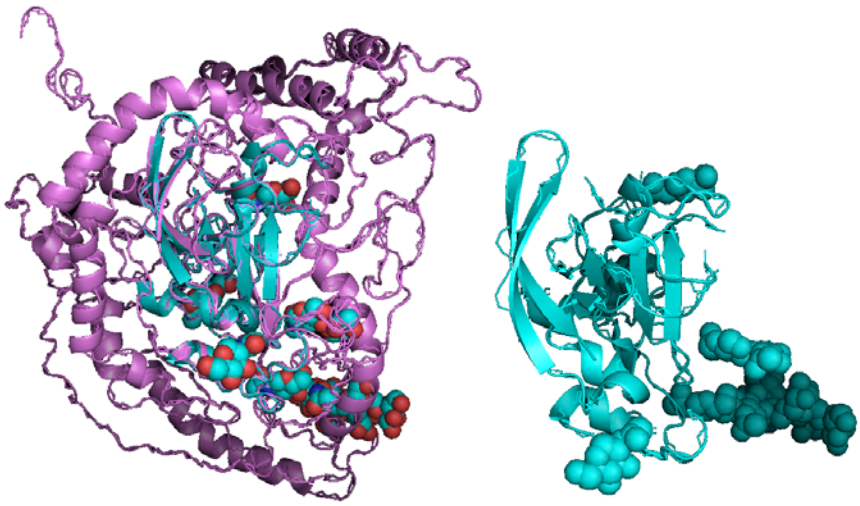
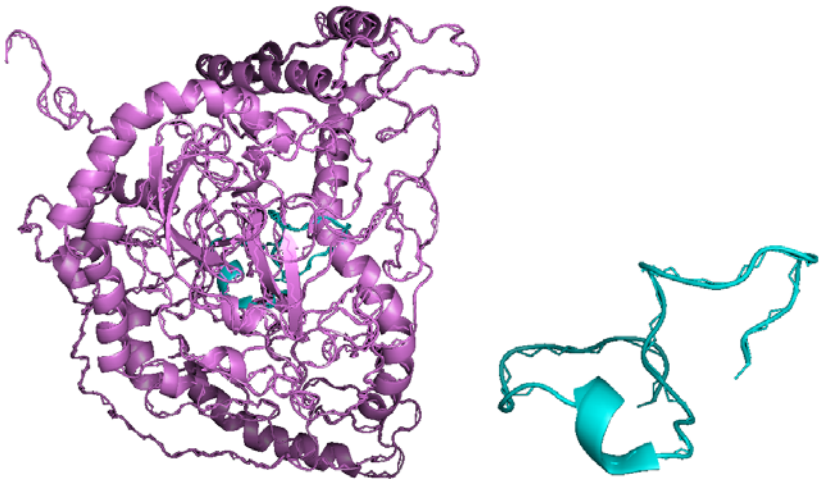


Figure 3 Visualisation of B cell conformational epitope of consensus CE1E2 protein. A total of 29 conformational epitopes are indicated by red color. N-terminal is on midle-left (Met-1) and C-terminal on upper left (Ala-748).



(a)



(b)

Figure 4 Visualisation of consensus CE1E2 protein position to natural protein. Pictures on left side are superimpose of three dimensional consensus CE1E2 protein (purple) position to natural E2 protein, PDB ID. 4MWF:C (a) and to natural C protein, PDB ID. 1XCT:P (b).

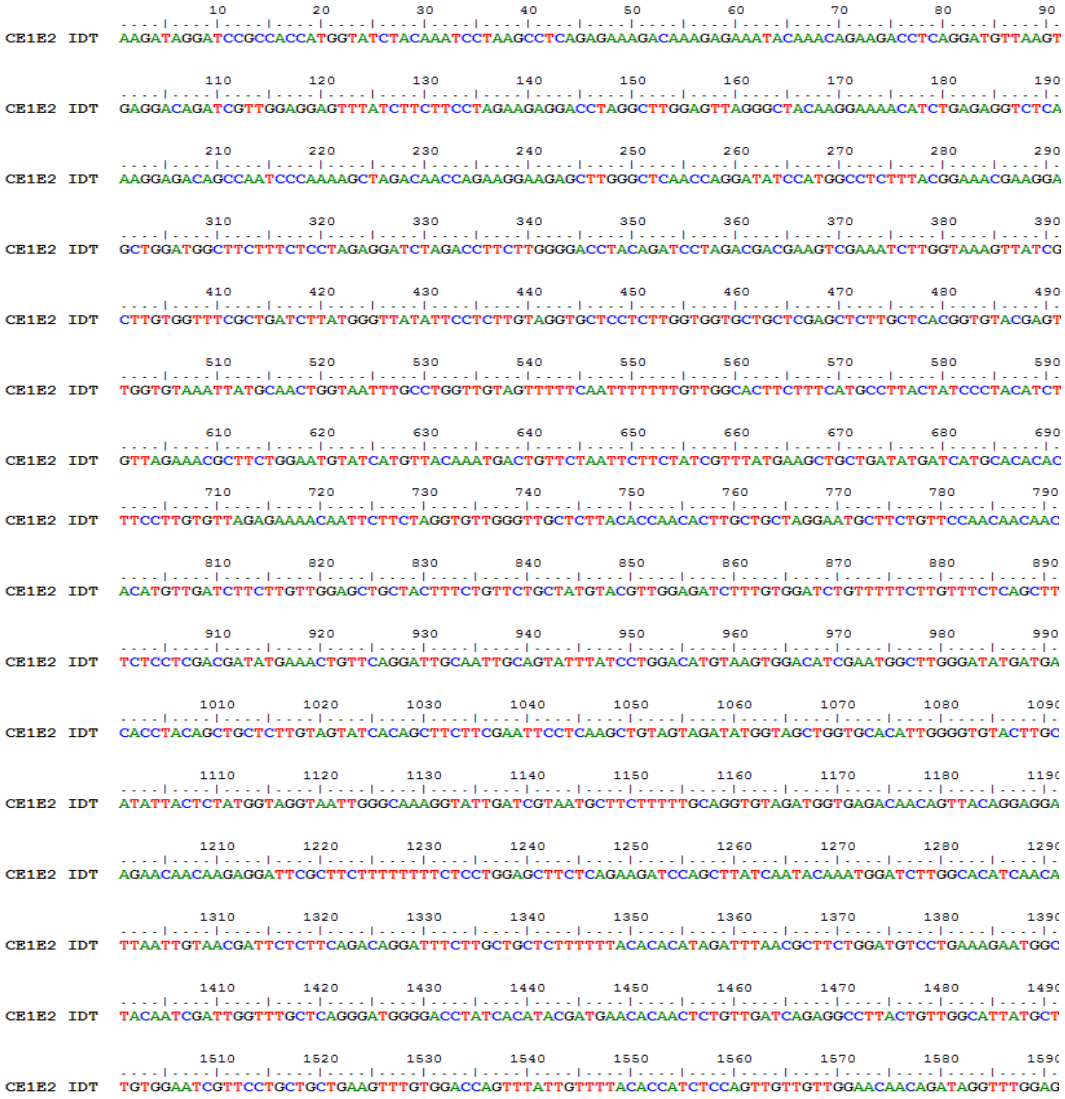


Figure 5 Nucleotide sequence of encoded gene of consensus CE1E2 protein. Consensus amino acid sequence of CE1E2 protein is translated by using CHO codon preference.

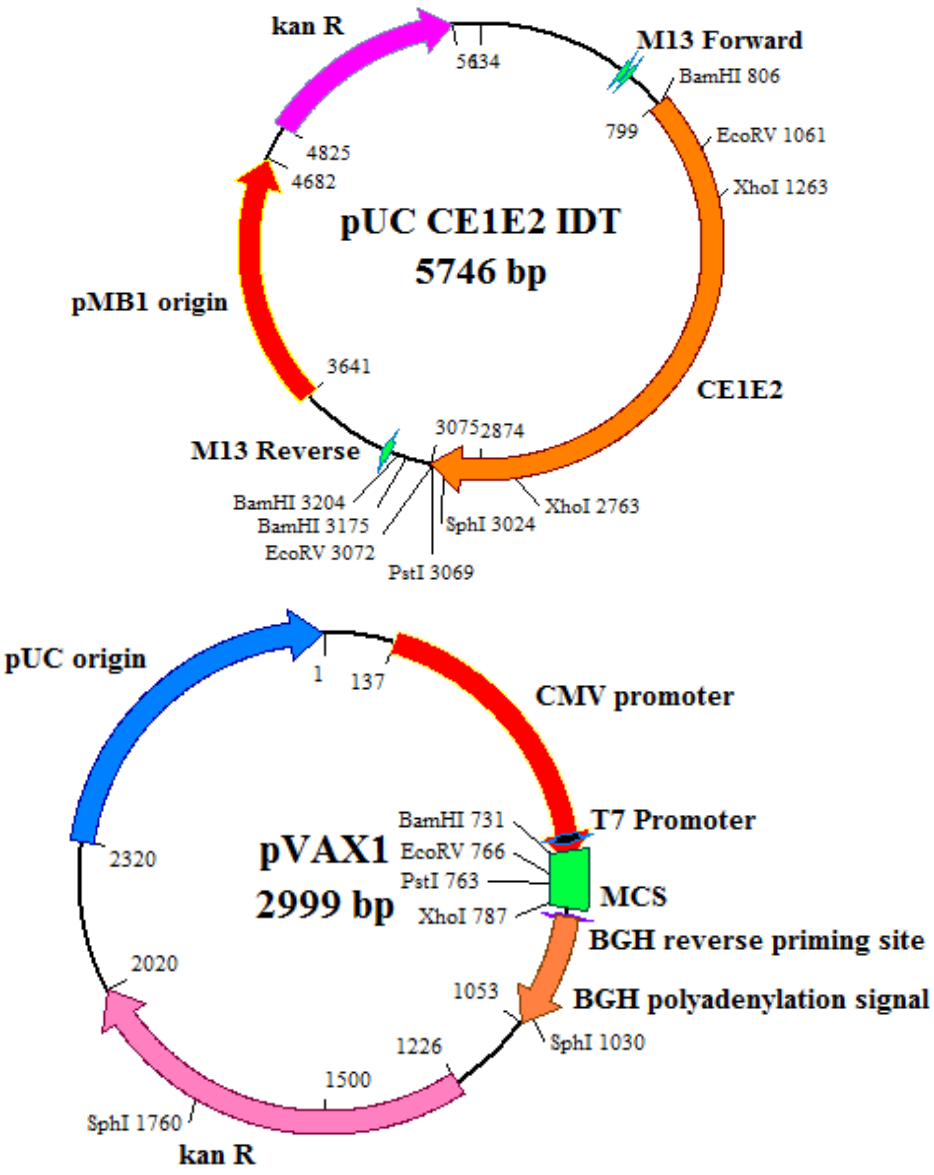


Figure 6 Vector map for pUC-CE1E2 and pVAX1©. Recombinant vector, pUC-CE1E2 has encoded gene of kanamycin resistance, universal priming site M13 and *BamH*I and *Pst*I restriction sites. Expression vector, pVAX1© has encoded gene of kanamycin resistance, universal CMV and T7 forward and BGH reverse priming site and *BamH*I and *Pst*I restriction sites.

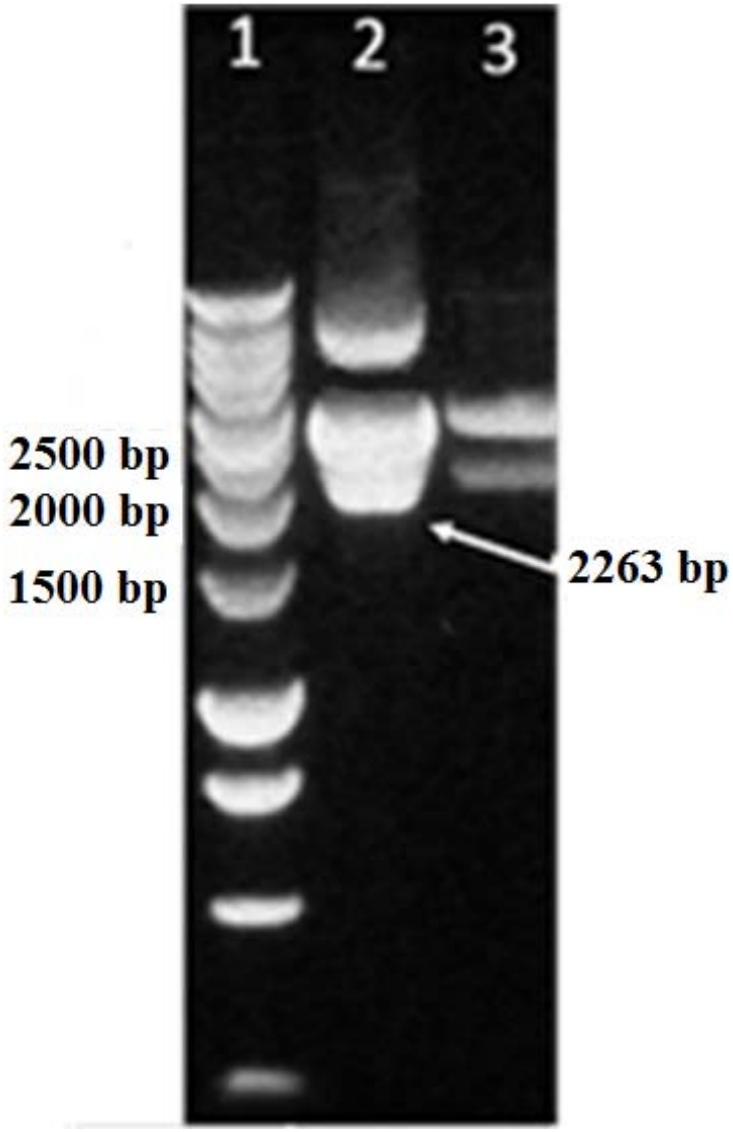


Figure 7 Electroforegram of pUC-CE1E2 and pVAX1© that have been digested by restriction enzyme. Lane 1, 1 kb marker; lane 2, digested pUC-CE1E2, sized 2000 bp, 3000 bp, and 5000 bp; lane 3, digested pVAX1©, sized 3000 bp.

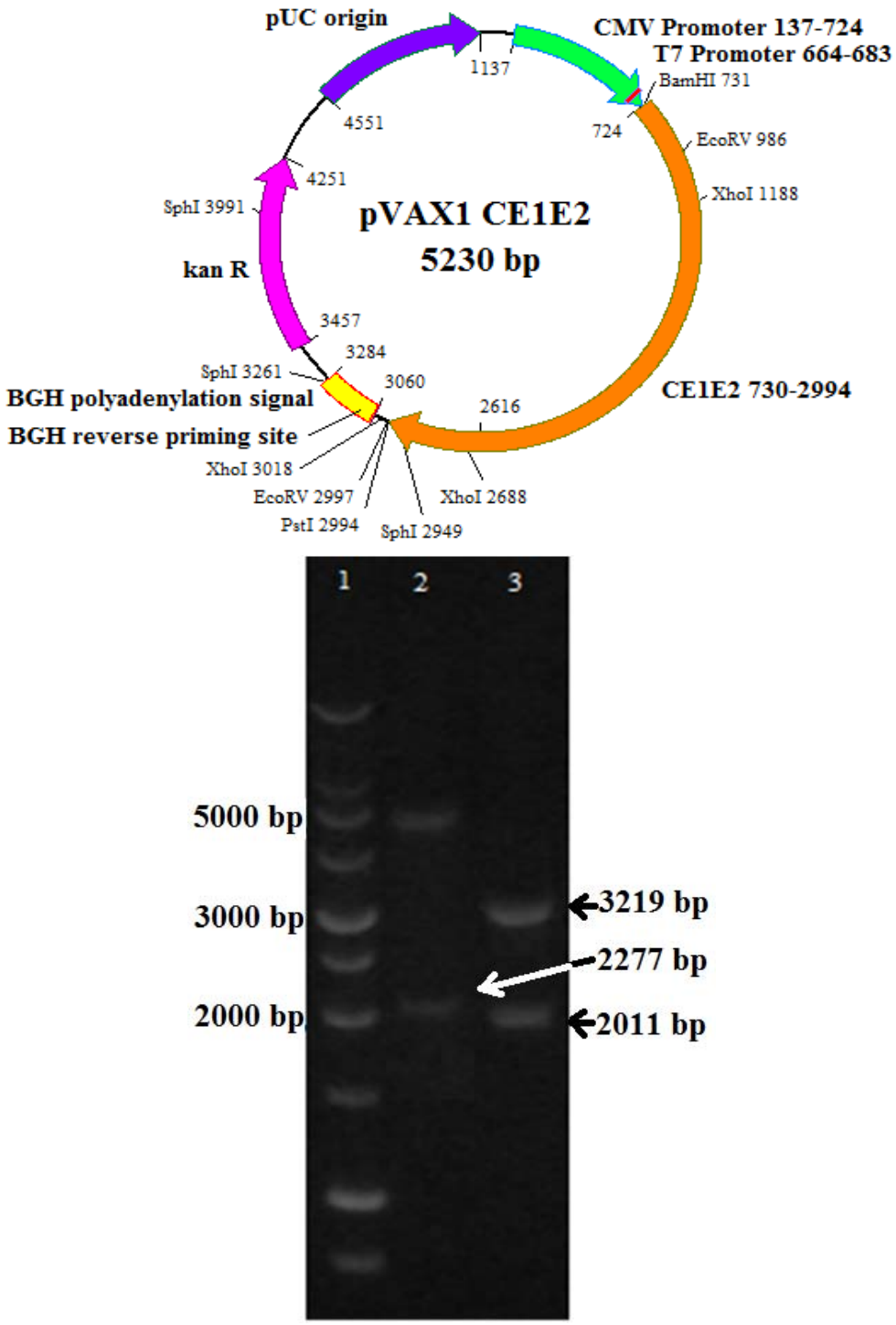
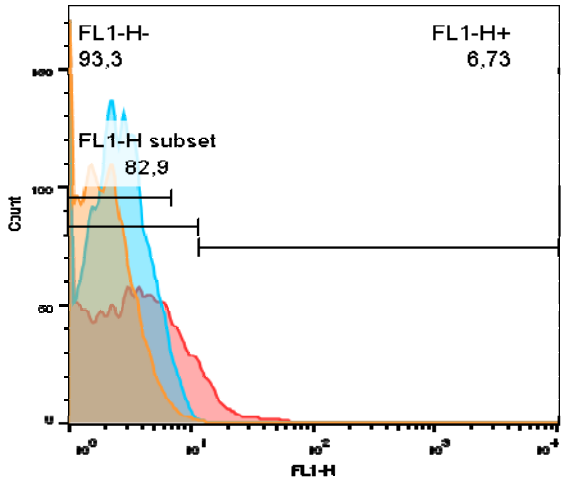
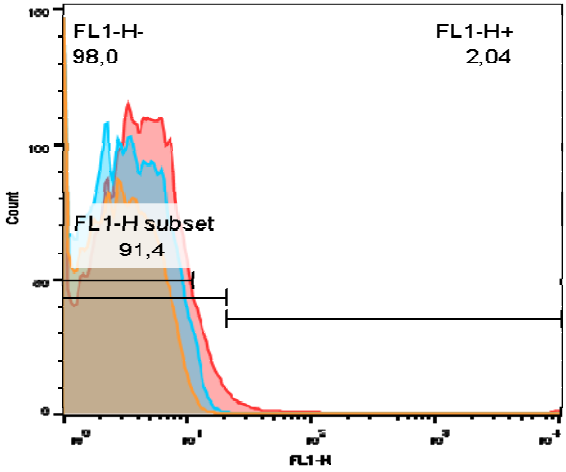


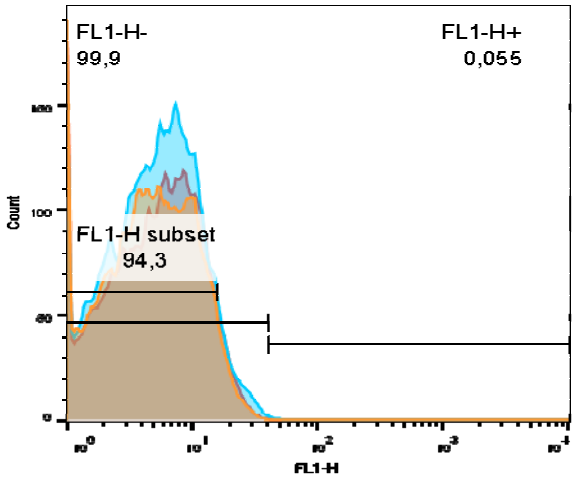
Figure 8 Vector map for pVAX1©-CE1E2 and electroforegram of pVAX1©-CE1E2 analysis. Lane 1, 1 kb marker; lane 2, amplification of encoded gene of CE1E2 protein from pVAX1©-CE1E2; lane 3, digested pVAX1© by *Eco*RV restriction enzyme.



(a)

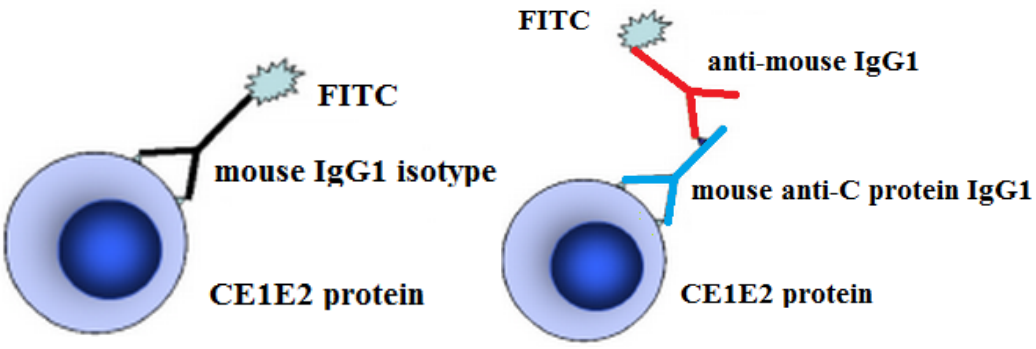


(b)



(c)

Figure 9 Observation of expression level of CE1E2 protein. Percentage of expression level on 24 hours, 6.73%; percentage of expression level on 48 hours, 2.04%; no significant expression level on 72 hours.



(a) (b)

Figure 10 Illustration of intra cellular staining of CE1E2 protein in CHO K1 cells. Direct staining of negative control (a) and indirect staining CE1E2 protein (b).

Additional files

Additional file 1 – as SeaView version 4.4.2

Figure 1 Phylogenetic tree of amino acid sequence of CE1E2 protein based on C protein. A total of 4.188 amino acid sequences were multiple aligned based on C protein. A consensus of amino acid sequence indicated by number 55552.
Binaries and full source code available from http://pbil.univ-lyon1.fr/software/seaview.html.

Additional file 2 – as FAS (BioEdit version 7.2.5)

Multiple sequence alignment result of 594 amino acid sequences of HCV genotype 1b publically available online.
Binaries and full source code available from
ftp://iubio.bio.indiana.edu/molbio/seqpup/

Additional file 3 – as FAS (BioEdit version 7.2.5)

4188 amino acid sequences of C protein HCV for phylogenetic tree input sequences, publically available online.