J Biochem Tech (2012) 3(4): 398-401
ISSN: 0974-2328

Apperception of conserved residues in variable hepatitis C virus E2 protein and human CD81 interaction

Venkatesh P, Radhika A, Snehal E

Received: 17 November 2011 / Received in revised form: 07 May 2012, Accepted: 23 June 2012, Published online: 27 October 2012
© Sevas Educational Society 2008-2012

Abstract

Hepatitis C Virus (HCV) has become a very prevalent disease affecting over 200million people worldwide. HCV amongst its various mechanisms of intrusion, enters the human cells by interaction with human CD81. The tetraspanin membrane protein CD81 interacts with envelope protein E2 of HCV. This helps viral entry into the host and facilitates further proliferation of the virus. Targeting this interaction could prove fruitful in eradicating the proliferation of the virus. We have *insilico*studied the interaction of E2 and CD81. Seventeen different sequences of E2 protein were homology modelled using 2ZCH_P - Chain P, Crystal Structure of Human Prostate Specific Antigen, as a template and using modeller 9v8 software. CD81 – E2 protein interaction was studied using Vakser Lab - GRAMM-X Protein-Protein Docking Web Server v.1.2.0. 170 models were generated, 10 each of a particular sequence. In almost all 17 protein models, residues cystine (2-6), Threonine (5-9) and Arginine (5-19) were found to be active in interactions with CD81.

Keywords: HCV, CD81, E2 protein, *insilico*, modeller, GRAMMX

Introduction

Hepatitis C is an infectious disease affecting the liver. It is caused by Hepatitis C Virus, belonging to the family of *flaviviridae*(Choo et al, 1989). HCV is an enveloped protein with a single stranded RNA genome. The exact mechanism(s) by which HCV enters and establishes in the body is not completely known. It is shown that HCV Envelope E2 protein interacts with CD81 (Pileri et al, 1998), which is believed to be the mode of viral entry into the cell. CD81 is considered to be the receptor for HCV (Cormier et al, 2004).

Different E2 protein sequences have been reported in NCBI. This shows that the virus has mutated itself over the period of evolution.

Venkatesh P*, Radhika A, Snehal E

Department of Molecular Biology, Sandor Proteomics Pvt Ltd, Hyderabad

*E mail: venkatesh@sandor.co.in

In spite of its evolutionary changes, the virus must have retained its structurally and functionally important amino acid residues which are responsible for its interactions with CD81 receptor.

Sequence Retrieval

The FASTA sequences of 18 different HCV E2 proteins reported in India, have been retrieved from NCBI with the following accession IDs (Table 1)

Table 1: Retrieved E2 proteins with their NCBI accession ID and the ID used in the rest of the paper

E2 protein Accession ID	ID (used in rest of the paper)
ABA08212.1	1
ABA08215.1	2
ABA08216.1	3
ABA08217.1	4
ABA08198.1	5
ABA08199.1	6
ABA08200.1	7
ABA08201.1	8
ABA08202.1	9
ABA08203.1	10
ABA08211.1	11
ABA08205.1	12
ABA08206.1	13
ABA08207.1	14
ABA08208.1	15
ABA08209.1	16
ABA08213.1	17

3D Modelling using Modeller9v8

The MODELLER software (Sali et al, 1993); a program for comparative protein structure modeling optimally satisfying spatial restraints derived from the alignment and expressed as probability density functions (pdfs) for the features restrained. The pdfs restrain C^{α}–C^{α} distances, main-chain N–O distances, main-chain and side-chain dihedral angles. The 3D model of a protein is obtained by optimization of the molecular pdf such that the model violates the input restraints as little as possible. The molecular pdf is derived as a combination of pdfs restraining individual spatial features of the whole molecule. The optimization procedure is a variable target function method that applies the conjugate gradients algorithm to positions of all non-hydrogen atoms.

The query sequence from Homo sapiens was searched to find out a template by the BLAST (Basic Local Alignment Search Tool) (Altschul et al, 1997) program against PDB (Protein Databank) (Bernstein et al, 1977). Sequences that showed maximum similarity with high score and less *E*-value were aligned and were used as a reference structure to build a 3D model for E2 proteins. The co-ordinates for the structurally conserved regions (SCRs) for E2 protein were assigned from the template using multiple sequence alignment, based on the Needleman-Wunsch algorithm (Needleman et al, 1970). The various options used for generating the model are discussed in table 2. 100 models were generated of which the least energy model was picked.

Table 2: parameters used in running modeller software

INCLUDE	#	Include the predefined TOP routines.
SET ALNFILE	=	'alignment.ali'
SET KNOWNS	=	'2F3O'
SET SEQUENCE	=	'1abc'
SET HETATM_IO	=	Off
SET WATR_IO	=	Off
SET HYDROGEN	=	Off
SET STARTING_MODEL	=	1
SET ENDING_MODEL	=	100

Structure assessment

The generated models were assessed for quality. The models were checked for φ and ψ torsion angles of Ramchandran plots with the help of PROCHECK (Laskoswki et al, 1993). The ramachandran plot is studied for residues in most favoured regions, additional allowed regions, generously allowed regions and disallowed regions. On obtaining a quality model, it is used for interaction studies with CD81 (PDB ID:1G8Q).

Protein – Protein interactions studies

Protein protein interactions were studied between E2 protein models generated, with Human CD81 individually. Vasker Lab- GrammX Protein – Protein Docking Web Server v.1.2.0 (Tovchigrechko et al, 2006) is used for studying the protein protein interaction studies.

Active site residues

The active site residues were studied using Swiss PDB Viewer (Guex et al, 1997). Using the various options available in spdbv, the residues sharing hydrogen bond interactions between CD81 and E2 protein are studied.

Results and Discussion

The sequences were retrieved from NCBI and blasted for template search.2ZCH_P - Chain P, Crystal Structure of Human Prostate Specific Antigen Complexed with an Activating Antibody was found to be the template for all the 17 different models. The percentage of similarity varied which is shown in table 3.

Tertiary Structures of proteins

The proteins were homology modelled using the modeller software (Sali et al, 1993). The models were tested for quality using procheck (Laskoswki et al, 1993). The number of Helixes, Strands and Turns of the 17 models are shown in table 4.

Table 3: The template and similarity percent with respective E2 proteins.

ID	Template	Similarity (%)
1	2ZCH_P	93
2	2ZCH_P	96
3	2ZCH_P	88
5	2ZCH_P	96
6	2ZCH_P	96
7	2ZCH_P	93
8	2ZCH_P	91
9	2ZCH_P	96
10	2ZCH_P	96
11	2ZCH_P	96
12	2ZCH_P	95
13	2ZCH_P	94
14	2ZCH_P	93
15	2ZCH_P	85
16	2ZCH_P	93
17	2ZCH_P	91

Table 4: The number of helixes, strands and turns for the 17 different E2 proteins

ID	Helix	Strand	Turn
1	2	6	6
2	2	6	7
3	2	6	6
4	2	6	6
5	1	6	6
6	2	6	6
7	2	6	6
8	2	6	6
9	2	6	6
10	2	6	6
11	2	6	6
12	2	6	6
13	2	6	6
14	2	6	7
15	2	6	6
16	2	6	6
17	2	6	6

The ramachandran plots of the 17 proteins were obtained. The residues in most favoured regions, additional allowed regions, generously allowed regions and disallowed regions of each protein is shown in table 5.

The 3 dimensional views of the 17 proteins are shown in the below figure 1. All proteins look structurally intact with minor changes.

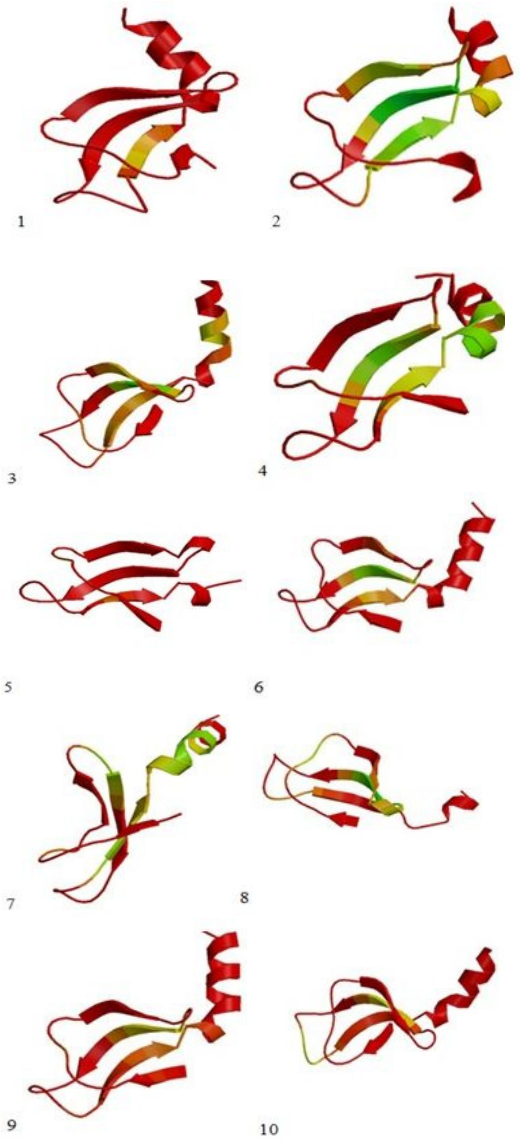
The protein - protein interactions provided 170 models, 10 models each of a particular E2 protein interacting with CD81. The active site residues of all the 17 E2 proteins were studied. One suchinteraction of ABA08198.1 with CD81 is shown in the Figure 2. The active site residues are highlighted in the clustalW (Thompson et al, 1994) analysis shown below.

Even though various amino acids are evolutionarily conserved, only few are functionally significant. It can be observed from the above clustalW image that rtesiduesCystine (2-6), Threonine (5-9) and Arginine (5-19) are the predominant residues. These residues are conserved over the years and could be playing a crucial role in interacting with CD81. The other conserved residues could be of structural significance.

Mutation based studies of these amino acid residues could reveal noteworthy results. These residues, if found important can be potential drug targets.

Table 5: Ramachandran plot analysis of the different 17 E2 proteins.

ID	Residues in most favoured regions	Residues in additional allowed regions	Residues in generously allowed regions	Residues in disallowed regions
1	81.5	14.8	3.7	0
2	85.7	12.2	2.0	0
3	87.5	12.5	0	0
4	87.5	10.4	2.1	0
5	74.5	23.4	2.1	0
6	86.0	14.0	0	0
7	85.7	14.3	0	0
8	80.4	19.6	0	0
9	86.3	11.8	2	0
10	89.1	10.9	0	0
11	74.0	26.0	0	0
12	91.3	6.5	2.2	0
13	84.4	15.6	0	0
14	76.4	20.0	3.6	0
15	81.6	18.4	0	0
16	89.6	10.4	0	0
17	89.6	10.4	0	0



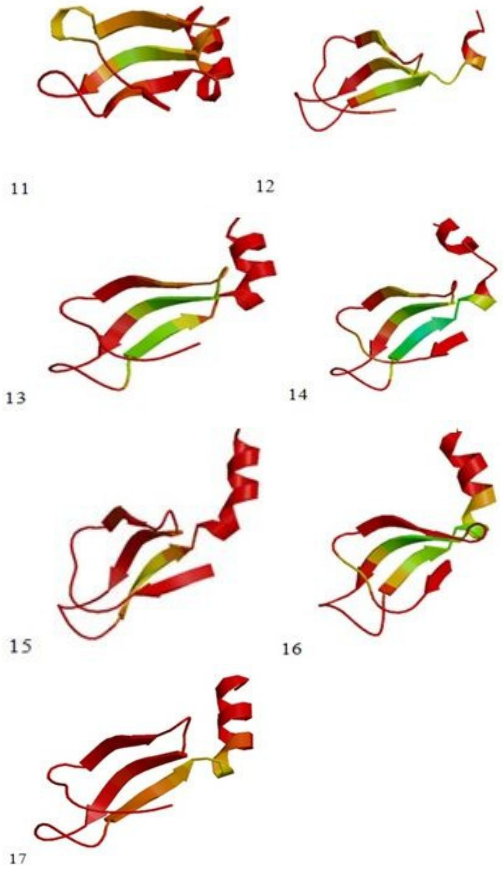


Figure 1: 17 different E2 protein models generated using modeller 9v8, with 2ZCH_P - Chain P, Crystal Structure of Human Prostate Specific Antigen as template.

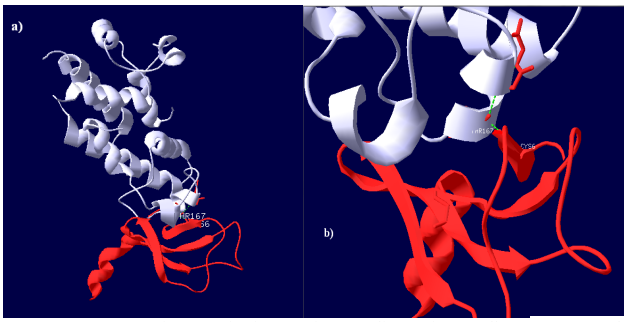


Figure 2: a) Protein protein interaction studies using spdbv (Guex et al, 1997). White ribbon protein – CD81 interacting with red ribbon protein – E2 protein. b) closeup of interacting aminoacids Thr167 and cys6, along with hydrogen bond (green dotted lines)

Conclusion

HCV virus has undergone conformational changes over the years for its existence. Different E2 proteins have been reported with varied sequences. From our study, we conclude that three amino acids cystine, thrionine and arginine, could be of prime importance for the virus to interact with the host. The other parts of the sequence have undergone changes over the period of evolution. This might affect the virus structurally but not functionally.

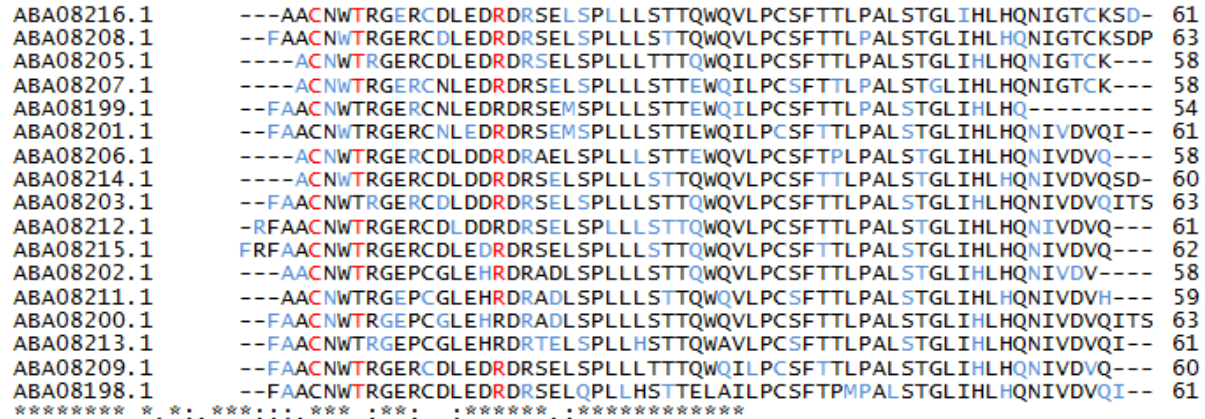


Figure 3: ClustalW (Thompson et al, 1994) analysis of 17 E2 proteins. Proteins observed to be involved in protein protein interactions have been highlighted. The residues highlighted in red are expected to be significant as they are active in all the proteins

Further, if these amino acids are mutated and the virus is allowed to interact with the host, it can be found and confirmed that these amino acids are the functional amino acids which are helping the viral entry and growth. Drug targets can be created on these particular amino acids.

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402

Bernstein F C, Koetzle T F, Williams G J B, Meyer E F Jr, Brice M D, Rogers J R, Kennard O, Shimanouchi T &Tasumi M (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.*,112:535-542

Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M (1989) Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science, 244:359-362

Cormier, E., F. Tsamis, F. Kajumo, R. Durso, J. Gardner, and T. Dragic. (2004) CD81 is an entry coreceptor for hepatitis C virus. Proc. Natl. Acad. Sci.USA 101:7270–7274

Guex, N. and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18:2714-2723

Laskoswki, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 26:283-291

Needleman, S. B. and Wunsch, C. D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48:443-453

Pileri, P., Y. Uematsu, S. Campagnoli, G. Galli, F. Falugi, R.Petracca, A.J. Weiner, M. Houghton, D. Rosa, G. Grandi, and S. Abrignani. (1998) Binding of hepatitis C virus to CD81.*Science.*282:938–941

Sali.A& Blundell T L. (1993) Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234:779-815

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. 22(22):4673-80

Tovchigrechko A, Vakser IA (2006) GRAMM-X public web server for protein-protein docking. Nucleic Acids Res. 34:W310-4