J. Product. & Dev., 22(1): 83-102(2017)

3-D STRUCTURE PREDICTION AND ANALYSIS OF THE P7-TRANSACTIVATED PROTEIN1OF HEPATITIS C VIRUS

Mahmoud M. El Hefnawi[1*]; Mohamed E. Hasan[2]; Amal Mahmoud[2];Yehia A. Khidr[3]; El-Sayed A. El-Absawy[2] and Alaa A. Hemeida[2]

[1] *Informatics and Systems Department, Division of Engineering Research Sciences, the National Research Centre, Egypt.*

[2] *Bioinformatics Department, Genetic Engineering and Biotechnology Research Institute, Sadat City University, Egypt.*

[3] *Plant Biotechnology Department, Genetic Engineering and Biotechnology Research Institute, Sadat City University, Egypt.*

ABSTRACT

The p7-transactivated protein of Hepatitis C virus is a small integral membrane protein of 127 amino acids, which is crucial for assembly and release of infectious virions. Ab initio and comparative modelling, is an essential tool to solve the problem of protein structure prediction and to comprehend the physicochemical fundemental of how proteins fold in nature. Only one domain (1-127) of p7 had been predicted using the systematic in silico approach, Threa Dom. I-TASSERwas ranked as the best server for full-length 3-Dprotein structural predictions of p7 where the benchmarked scoring system such as C-score, TM-score, RMSD and Z-score are used to obtain quantitative assessments of the I-TASSER models. Scanning protein motif databases, along with secondary and surface accessibility predictions integrated with post translational modification sites (PTMs) prediction revealed functional and protein binding motifs. Three protein binding motifs (two Asp/Glutamnse, CTNNB1-bd_N) with high sequence conservation and two PTMs prediction: Camp_phospho_site and Myristyl site were predicted using BLOCKS and PROSITE scan.

These motifs and PTMs were related to the function of p7 protein in inducing ion channel/pore and release of infectious virions. Using SCOP, only one hit matched protein sequence at 71-120 andwas classified as small proteins and FYVE/PHD zinc finger super family.

Integrating this information about the p7protein with SCOP and CATH annotations of the templates facilitate the assignment of structure–function/ evolution relationships to known and newly determined protein structures.

Keywords: Hepatitis C virus, p7 protein, *Ab initio*, proteinstructure prediction, motifs and PTMs

.

INTRODUCTION

Hepatitis C virus (HCV) infection is a major cause of chronic liver disease. It is considered as a main threat to global public health. According to the WHO, (more than 185 million people) is chronically infected with HCV around the world and 350 000–500 000 deaths estimated annually (Graham and Swan, 2015; Gower *et al*., 2014), because of chronicity, hepatic fibrosis, cirrhosis, and, increasingly, hepatocellular carcinoma (HCC) (Sugiyama, 2004).

HCV genome consists of a 9.6-kb positive strand RNA virus belonging to the genus *Hepacivirus,* family *Flaviviridae*(Simmonds, 2013).It encodes a single polyprotein precursor which processed by both host and viral proteases into ten mature proteins (Niepmann, 2013; Moradpour and Penin, 2013). The **core** protein, which composes the nucleocapsid; the viral envelopeE1 and E2 glycoproteins; the p7 viroporin required for virus assembly; and the replication machinery consisting of NS3, NS4A, NS4B, NS5A, and NS5B (Moradpour and Penin, 2013).

The p7-transactivated protein1 is a small integral membrane protein of 127 amino acids containing viral proteins from several virus families. P7 is crucial for manipulating membrane permeability for ions by forming cation-selective pores (Montserret *et al*., 2010)and for assembly and release of infectious **virions** (Gower *et al*., 2014).

One of the protein structure prediction software applications is the prediction of three dimensional structures of virus' proteins, to elucidate their functions. This will aid in drug design, and integrative understanding of viral processes. The process of prediction of the three dimensional structure of a protein from its amino acid sequence are divided into three categories; Comparative modeling (or Homology modeling), Fold recognition (or Threading), and Ab-initio prediction (Free modeling). Prediction of unknown

structure based on known structure (template), in which the sequence similarity is more than 30% defined as homology modeling (Cheng and Baldi, 2007; Wu and Zhang 2008). The fold recognition, which utilizeda library of templates, or *ab-initio* methods are used to predict 3D for domains that have no homologue.

Ab-initio methods can be used to predict the protein structure from the sequence information when no suitable structure templates can be found (Klepeis *et al*., 2005 and Liwo *et al*., 2005). Free modeling is the 'Holy Grail' of protein structure prediction because successes in building correct topologies (3~6Å) for small proteins (<150 residues) (Wu and Zhang., 2008).

In this study, systematic *in silico* structural and functional analysis of HCV p7-transactivated protein was used to predict the 3-D structure and to better understand structural and functional of P7 protein.

Methods

The methodology to analyze the p7 protein includes the prediction of conserved regions, domains, secondary structures, three-dimensional structures, post-translational modification sites, signatures and motifs.

Conserved regions

Multiple sequence alignments of the p7 protein were conducted from different genotypes using the PROMALS server (Pei and Grishin, 2007), Clustal Omega server (Sievers *et al.,* 2011) and the BIOEDIT software (Hall, 2011) was used to extract the conserved regions. The PROMALS is up to 30% more accurate compared to the best alignment methods with improvement for distantly related sequences. Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between three or more sequences. The BIOEDIT software is a user friendly biological sequence alignment editor and it is intended to provide basic functions for protein sequence editing, alignment, manipulation and analysis.

Molecularevolutionary and phylogenetic analysis

The evolutionary history was inferred using the Neighbor Joining method. The process of finding a phylogenetic tree using maximum likelihood involves finding the topology and branch lengths of the tree that will give us the greatest probability of observing amino acid sequences in our

data. So, for Phylogenetic analysis Mafft server (Katoh *et al.*,2015), Clustal Omega server and MEGA7 (Kumar *et al.*,2016) software were used.

Domain Separation

The first step for the three-dimensional structure prediction, is to accurately separate the domains. The domains of p7 had been predicted using one of the top domain prediction servers in CASP8, CASP9 and CASP10, Threa Dom (Xue *et al.*, 2013). This server is a template-based algorithm for protein domain boundary prediction. The domains are further subdivided into sub-domains using ProDom (Catherine *et al.*, 2005), which is a comprehensive database of protein domain families generated from the global comparison of all available protein sequences.

Secondary structure and solvent accessibility prediction

To maximize the accuracy of prediction, consensus results from different prediction servers are used. Seven servers have been used for prediction; Porter (Pollastri and McLysaght, 2005), PSIPRED (McGuffin *et al.*, 2000), SSPRO [Cheng *et al.*, 2005], Yaspin [Lin *et al.*, 2005], NPS@ SOPMA [Combet *et al.*, 2000], JPred [Drozdetskiy *et al.*, 2015] and PredictProtein (Rost *et al.*, 2004). All these servers are based on neural networks in their techniques. The Porter server uses the bidirectional recurrent neural networks (BRNNs) while PSIPRED uses two feed-forward neural networks. Predict Protein is a meta-service for sequence analysis that has been predicting structural (secondary structure, solvent accessibility, trans membrane helices, strands, coiled-coil regions, disulfide bonds and disordered regions) and functional features of proteins. Jpred4 provide predictions by the **Jnet** algorithm, one of the most accurate methods for secondary structure prediction, while Spro uses one-dimensional recursive neural network (1D-RNN) architecture. Also, YASPIN is a HNN (Hidden Neural Network) that uses the PSI-BLAST algorithm to produce a PSSM for the input sequence, which it then uses to perform its prediction.

Also, predict protein server was used to find the exposed and buried regions for p7 protein. This server classifies each amino acid as being in one of 4 classes (all-alpha, all-beta, alpha-beta and mixed all others).

3-D Structure Prediction

The Galaxy WEB (*Koet al.*,2012), Swiss-Model (Biasini *et al.*,2014) and I-TASSER (Yang *et al.*,2015) are the servers that are used for homology

modeling prediction. For domains that don't have a direct homologue, fold recognition prediction meta servers 3D-Jury (Ginalski *et al.*,2003), LOMETS [Wu S. and Zhang, 2007] and Pcons (Wallner and Elofsson 2005) are used. For *ab-initio* method the I-TASSER, Phyre2 (Kelley *et al.*,2003) and QUARK (Xu and Zhang, 2012) server which are the top servers in the free modeling in CASP9 and CASP10 are used. For instance, I-TASSER (as 'Zhang-Server') was ranked as the No 1 server for protein structure prediction in recent community-wide CASP7, CASP8, CASP9, CASP10, and CASP11 experiments. The I-TASSER server is in active development with the goal to provide the most accurate structural and function predictions using state-of-the-art algorithms. Structural templates are first identified from the PDB by multiple threading approach LOMETS; full-length atomic models are then constructed by iterative template fragment assembly simulations. I-TASSER builds the initial full-length models by filling the gaps using a self-avoiding random walk of Cα–Cα bond vectors of variable lengths from 3.26 to 4.35 A˚.

Model refinement

Galaxy WEB, Modrefiner (Xu and Zhang, 2012) and $3D^{refine}$ (Bhattacharya and Cheng, 2012) servers are web services for consistent and computationally efficient protein structure refinement by energy minimization and molecular dynamics methods. The role of these servers is simultaneous improvement in both global and local structural qualities of the initial models to bring it closer to the native state in a computationally inexpensive manner. There finement process is based on two steps, first step is based on optimization of hydrogen bonding (HB) network and second step applies atomic-level energy minimization on the optimized model using a composite physics and knowledge-based force fields.

Model evaluation

Sampling and ranking structural models are the two major challenges of protein structure prediction. There is novel large-scale model quality assessment (QA) methods in conjunction with model clustering to rank and select protein structural models. We evaluated the refinement category predictions through different measures (GDT-TS, GDT-HA, TM-score, Z-score, Mol Probity (MP) score, Qmean score, estimated absolute model quality Z score, Clash score and Root Mean Square Deviation RMSD).These measures examine complementary aspects of model quality including overall

fold, as well as the distributions of inter-atomic contacts and dihedral angles. QMEAN Server for Model Quality Estimation (Benkert *et al.*, 2008), Swiss Model Workspace (Benkert *et al.*, 2011), PROSSES (Berjanskii *et al.*, 2010), SAVES, PROCHECK (Laskowski *et al.*, 1993), WHAT_CHECK (Hooft, 1996) and TM-align (Sigrist *et al.*, 2013)web servers were used for evaluation of the accuracy of automated protein structure prediction methods of p7 protein.

Functional motifs prediction

The principle advantage of motifs/fingerprints is that they can detect distant sequence relationship. Also, protein-protein interactions (PPI) can be discovered by deriving information about motifs. There are two approaches used to derive information about motifs, the first approach is using PROSITE (Henikoff *et al.*, 2000) web server which matches the regular expressions with a query sequence. The second approach is using BLOCKS (Henikoff *et al.*, 1999), PRINTS (Attwood *et al.*, 2015), PRODOM (Kahn *et al.*, 2008) and SMART (Letunic *et al.*, 2015) web servers that uses a statistical model, such as position-specific scoring matrices (PSSMs), profile, or Hidden Markov models (HMMs) to preserve the sequence information from a multiple sequence alignment and express it with probabilistic models. SMARTA rchitecture allows users to search for specific domain architectures using an AND/NOT logic. SMART analysis of a query sequence reveals not only domains, but also intrinsic features such as signal sequences, trans membrane helices, coiled coil regions and compositionally biased regions. Furthermore, Scan site (Wan *et al.*, 2008) identifies short protein sequence motifs which are represented as PSSM based on results from oriented peptide library and phage display experiments.

RESULTS AND DISCUSSION

The p7-transactivated protein 1 of HCV of 127 amino acids (accession number Q6PLM3) was retrieved from Uni Prot Knowledgebase (Uni Prot KB) and used in this study.

Molecular evolutionary and phylogenetic analysis

The evolutionary history was deduced using the Maximum Likelihood method based on the JTT matrix-based model. phylogenetic tree was established using Neighbor-Joining (NJ) method and phylogenetic analysis of

p7-transactivated protein 1 of HCV (Figure 1) revealed that HCVp7 protein was closely related to P7 protein of Simian pegivirus and putative p7 protein of Hepatitis GB virus B (Figure 1).

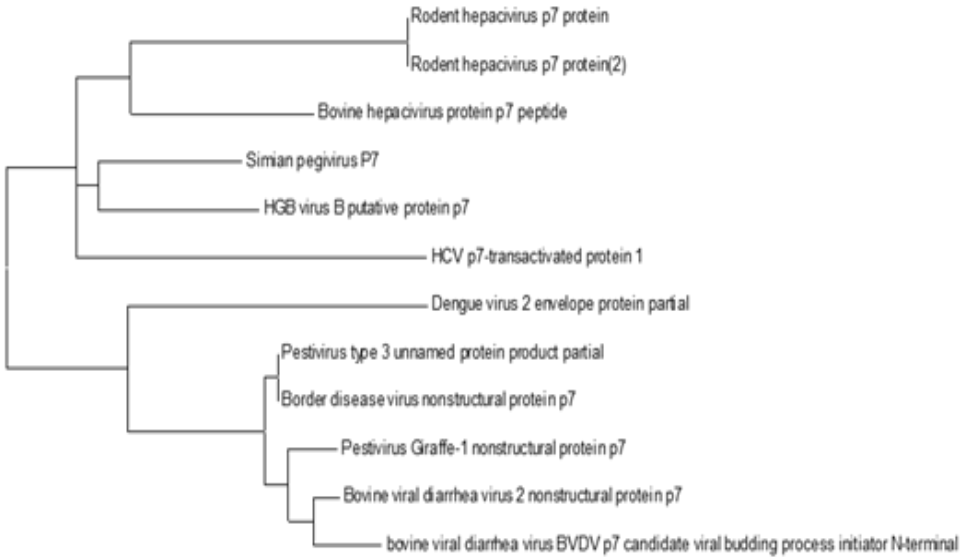


Figure (1): Molecular Phylogenetic analysis by Maximum Likelihood method.

Domain separation

Using Threa Dom and Pro Domp7-transactivated protein hadonly one domain (1-127) with Cutoff 0.66, Score 154 (63.9 bits), E value 3e-09 and Identities 47/127 (37%).

Secondary structure and solvent accessibility prediction

To increase the accuracy of secondary structure prediction, we combined many neural network predictions. Using a consensus result from the output of seven robust servers: Porter, Prof, PSI Pred, SSPro, SOPMA, Predict Protein and YASPIN secondary structure prediction of P7 protein composed of Alpha helix 7%, Beta strand 8%and random Coil 85%. Heliceswere specified at 88–92 and β-sheets at 16-19, 28-33, 41, 76-77, 93-94 and 97-100.Also, five disulfide bond were found at 17-31, 29-61, 92-94, 97-103 and 112-116 (Figure 3). كيف يكون قبل2 Fig. 3

Using predict protein server, the solvent accessibility of the predicted protein interaction motifs of the P7 protein were found to be mostly exposed 58%, intermediate 5% and buried 37%.In addition, 17 protein binding sites to which ligands may form a chemical bond were detected at 10,12,24,30,33,36,38-41,44-51,53-58, 67,71, 75-76, 85-94, 97, 100, 109-116 and 124 (Figure 3).

3-D Structure Prediction:

The 3-D model, initial models are constructed, refined, evaluated and finally the one with highest quality is chosen.

Construction of initial model using target-template alignment

The structural models for the unaligned regions, various servers such as I-TASSER, Phyre2 and QUARK were designed while Galaxy WEB, Swiss-Model and LOMETS were designated for the aligned regions. QUARK server constructed best ten correct protein 3D model from amino acid sequence based on estimated TM-score 0.2999 ±0.0764, while I-TASSER built top five models based on confidence score(C-score) of each model (equal to -3.12), which is typically in the range of (-5, 2). The TM-score 0.36±0.12 and RMSD11.5±4.5Å were estimated based on C-score and protein length. Galaxy WEB provided five models for each query sequence and select templates for modeling by rescoring HH search results. SWISS-MODEL gave three models based on QMEAN score (-1.25), while Phyre2 used advanced remote homology detection methods to build 3D models. LOMETS generated top 10 models based on C- score (medium) andZ-score (6.292) of the template using 9 locally-installed threading programs.

Reduced-level structure assembly and refinement simulations:

The refinement of P7 protein was the second step for structure prediction. The results from Galaxy WEB, Modre finer and 3-D^{refine}servers succeed to draw the initial starting models closer to their native state, in terms of hydrogen bonds, backbone topology and side-chain positioning. It also generates significant improvement in physical quality of global and local structures by reducing RMSD score (0.546) and increasing TM-score (0.9869) to initial model resulted from I-TASSER server.

Model evaluation and selection

Model evaluation was used to select the high quality 3-D model of correct fold from all the possible closest alternative conformations to the

native structure. The resulted scores from different evaluation servers such as QMEAN Server for Model Quality Estimation, Swiss Model Workspace, PROSESS, SAVES, TM-align, PROCHECK and WHAT_CHECK were summarized in Table 1. The best model was obtained from I-TASSER which was arranged as the best method in the server section of the recent 7th, 8th, 9th and 10th CASP experiment. I-TASSER server generated five full-length models with C-score (high), estimated TM-score (0.9977) and RMSD (2.07) for query p7 protein. A scoring function (C-score) has strong correlation with the TM-score for assessing the similarity of protein structure and estimating the accuracy of the I-TASSER predictions. The TM value of selected refined model (0.9977) was larger than 0.5, this meant that model was very accurate. Otherwise when a value is between 0.5 and 0.17 or less than 0.17, this give a model with a roughly correct topology or a random prediction regardless of the protein size (Zhang and Skolnick 2004). The QMEAN Z-score provides an estimate of the absolute quality of a model by relating it to reference structures solved by X-ray crystallography. The 'good' models (QMEAN Z-score equal to −0.34) which were resulted from I-TASSER, are depicted in green reach QMEAN Z-scores comparable to experimental structures (QMEAN Z-score ≥−0.65) whereas the 'medium' quality models are located in blue (0.65<QMEAN Z-score>−1.75). Bad models with large negative value(QMEAN Z-score ≤−2)are found in the dark red[37]. In addition to I-TASSER models had low value of RMSD 2.07Å which refer to high accuracy model (RMSD < 3Å) (Sitao and Yang 2007), Mol Probity score (2.054) and clash score (9.7). The best predicted three-dimensional structure of p7 protein by I-TASSER server is shown in (Table 1 and Figure 2).

The functionality of the top templates was ranked by MUSTER (Wu and Zhang 2008) and the best template structure is (1URK) plasminogen activator. A plasminogen activator is a serine protease of 130 residueswhich consist of α-helix 6%, β-sheet 11%. A plasminogen activator has two domains: plasminogen activator urokinase-type (44 residues, 6-49) and Urokinase-type plasminogen activator (86residues, 50-135). The amino-terminal fragment (ATF) of urokinase-type plasminogen activator is a two sub-domain protein which consists of a growth factor and a kringle domains. The later are believed to play a role in binding mediators (*e.g.*, membranes, other proteins or phospholipids), and in the regulation of proteolytic activity (Atkinson and Williams, 1990).

Table 1: Evaluation of 3-D structure model from selected servers for p7 protein structure prediction.

Servers / Score	I-TASSER	LOMETS	Quark	SWISS-MODEL	GALAXY WEB	PHYRE2
RMSD	2.07	2.72	5.09	3.72	4.72	3.29
TM-Score	0.9977	0.9997	0.9868	0.8869	0.9982	0.8148
GDT-TS	0.5615	0.5231	0.1308	**0.0750**	0724	0.1115
GDT-HA	**0.**9311	0.9744	0.9331	**0.9397**	0.9733	0.9762
Q-mean	0.404	0.374	0.417	0.441	0.239	0.558
Estimated absolute quality (Z score)	-0.34	-0.98	-3.57	-4.04	-5.54	-0.411
Z-score mean	0.710	- 0.423	0.541	-0.404	0.691	2.564
Z-score mean RMS	2.0531	1.734	1.855	1.704	2.093	2.564
MolProbity	2.054	2.661	2.649	1.511	2.456	1.777
Clash score	9.7	19.4	26.9	13.7	5.0	7.6
Aligned length	122	118	127	29	127	21

Motifs prediction

Motif analysis using blocks server.

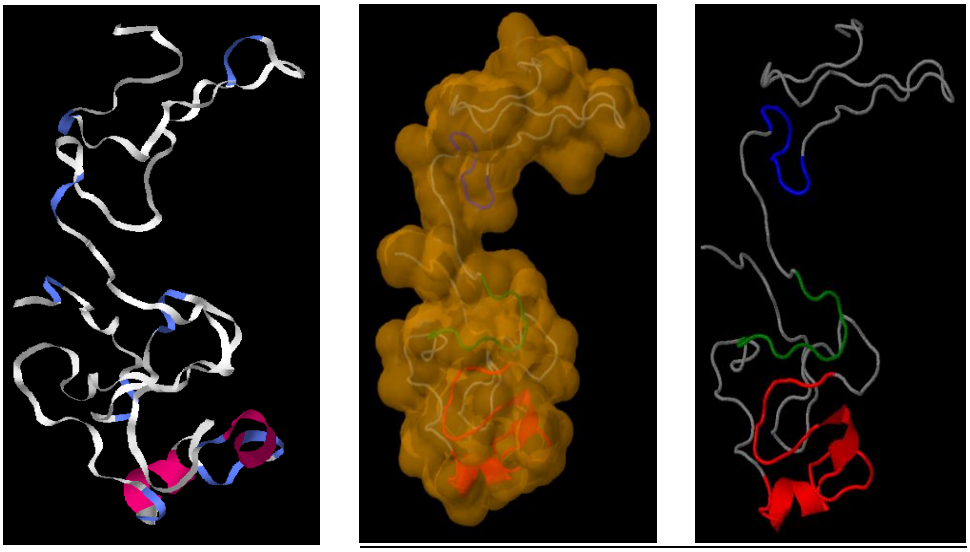
Motif analysis of HCV P7 protein using Blocks server wasshown on Table 2. Two blocks were detected in Asparaginase/glutaminase family with combined E-value 0.23, one of them (32-39) and other (71-100) has blocked E-value 10 and 44 respectively. One block (48-57) exists in the Beta-Catenine bound non-globular protein regions (CTNNB1 binding, N-teminal) family with combined E-value 0.25 and block E-value 0.29. In addition, only one block (51-95) was revealed in Anti-Mullerian hormone, N-terminal (AMH_N) family with combined E-value 0.94 and block E-value 0.84 as shown in (Table 2 and Figure 2).

Post-translational modification site prediction using Prosite server.

Using PROSITE database, different signatures were retrieved all over the p7- trans activated protein: camp_phospho_site at (115-118), (MYRISTYL)

Table (2). Motif analysis using Blocks server.

Block number	Family	Match Position	Strand	Blocks	Combined E-value	Block E-value
IPB006034 **IPB006034A** **IPB006034C**	Asparaginase /glutaminase	 32-39 71-100	1	2 of 5	0.23	 10 44
IPB013558 **IPB013558C**	CTNNB1 binding, N-teminal	48-57	1	1 of 7	0.25	0.29
IPB006799 **IPB006799D**	Anti-Mullerian hormone, N-terminal	 51-95	1	1 of 9	0.94	 0.84



(a) **(b)** **(c)**

Figure 2: The best predicted three-dimensional structure ofp7 protein by I-TASSER server. (a) Ribbon view of best predicted model, (b) Solvent-accessible surface view shows the exposed regions. (c)The cartoon view shows known and predicted motifs: The Asp/Glu **(IPB006034A, 32-39)** was highlighted in blue, CTNNB1-bd-N **(IPB013558C, 48-57)** was highlighted in green, Asp/ Gls **(IPB006034C, 71-100)** was highlighted in red and the rest of predicted model was highlighted in gray.

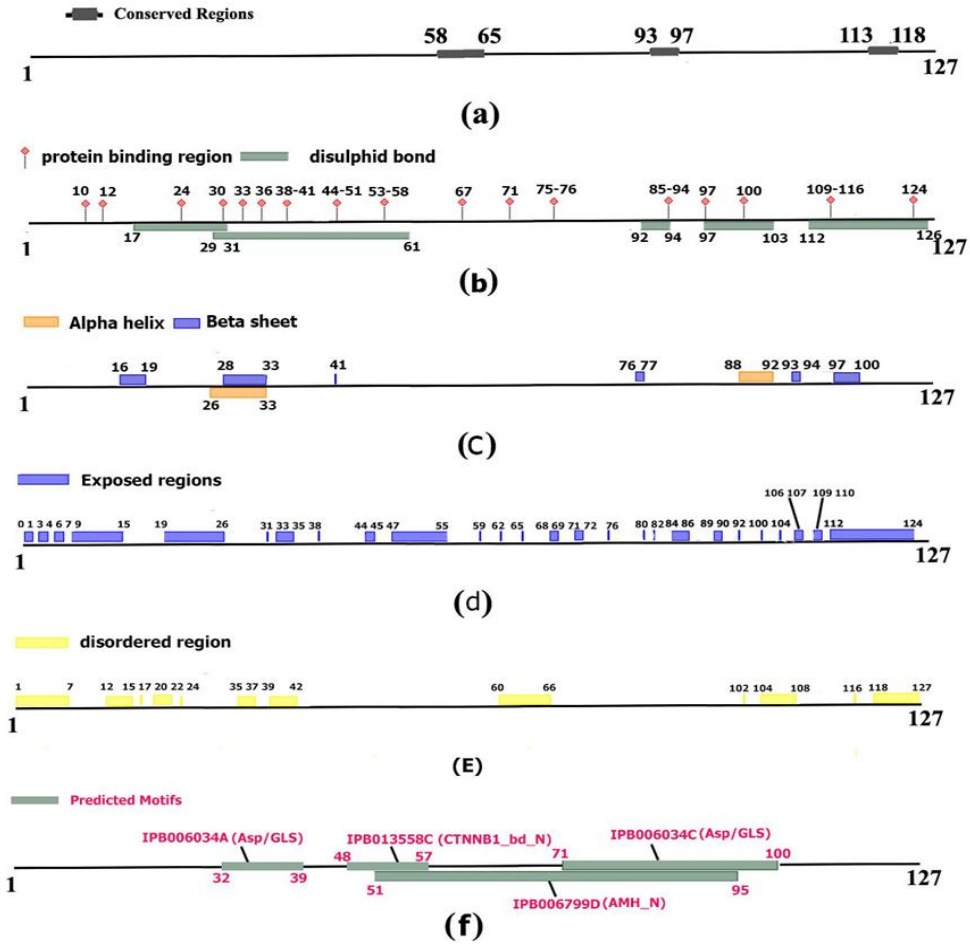


Figure 3: Integrative map of predicted results. In (a), the three predicted conserved regions between HCV family are specified at 58-65, 93-97 and 113-118. In (b) and (c), the consensus result of protein binding region from the seven secondary structure prediction servers, is specified. In (d), the exposed results from predict protein server are specified. In (e), the SMART results (intrinsic disorder regions) are specified. In (f), four important predicted motifs (blocks) are specified; IPB006034A(32-39aa), IPB00603AC (71-100), IPB013558C (48-57) and IPB006799D (51-95aa). Based on these motifs, the p7 protein is dispensable for RNA replication, but crucial for the production and release of infectious HCV particles from infected cells.

N-myristoylation site at (20-25, 35-40, 42-47, 80-85 and (104-109), Protein kinase C phosphorylation site (PKC_PHOSPHO_SITE) at (105-107, 114-116 and 123-125) and Proline rich region profile (PRO_RICH) at 48-72 (Table 3).

Table (3). Post-translational modification site prediction using Prosite server.

Category	**Signature**	**Matching positions**
RNA Associated Protein	Camp_Phospho_Site	115-118
Domain	Myristyl	20-25, 35-40, 42-47, 80-85 and104-109
Posttranslational Modifications	Pkc_Phospho_Site	105-107, 114-116 and 123-125
	PRO_RICH	48-72

Identification, annotation and analysis of domain architectures.

The SMART server allows the identification and annotation of genetically mobile domains and the analysis of domain architectures. For p7 trans activated protein sequence, the results from Predict Protein, Dis EMBL (Linding *et al.*, 2003) and SCOP servers had been returned to SMART and detected disorder region in proteins that are important for protein function. The specified disordered regions were located at 1-7, 12-15, 20-22, 24, 35-37, 39-42, 60-66, 102, 104-108, 116 and 118-127. Using SCOP, only hit matched protein sequence at 71-120 with *E*-value 1.30e-10, and it was classified as small proteins and FYVE/PHD zinc finger super family.

To facilitate the assignment of structure–function/ evolution relationships to both known and newly determined protein structures CATH and SCOP were used. The CATHEDRAL structure comparison algorithm was used to characterize structural diversity in CATH super families. Using CATH (v4.0), 67 matching structures were detected based on SSAP score (1-100), RMSD and length. The best hit (1sqgA03) matched p7 sequence 58 residues at 67-125 with SSAP score (62.2), RMSD (8.5) and classified this match as α/β proteins, 2-Layer Sandwich architecture, and α-β Plaits topology and sun protein; domain 3superfamily (Table 4).

The CATH website has been redesigned and now displays additional functional data, so functionally sun protein; domain 3 super family has been used to demonstrate the functionality of the Ribosomal RNA small subunit methyltransferase B.

Table 4. CATH Classification

Level	Description	Super family	Functional Family	Enzyme	Structural Cluster
Class	Alpha Beta	Sun protein; domain 3	Ribosomal RNA small subunit methyl transferase B	16S rRNA (cytosine (967)-C(5))-methyl transferase.	SSG 1
Architechure	2-Layer Sandwich				
Topology	Alpha-Beta Plaits				
Homology	Sun protein; domain 3				

Correlation of Structural and Functional Prediction

Secondary structure prediction of P7 protein consisted of α-helix 7%, β-strand 8% and random Coil 85%. From the consensus secondary structure and SMART results: The two Asp/Glu blocks includes α-helix at 88-92, while the third AMH_N block hit has α-helix at 88-92 and β-sheet at 76-77 and 93-94. The specified disordered regions and hot loops at 35-37 and 60-66 were located within Asp/Glu and AMH_Nblock hit, respectively. Investigating the interaction motifs of the P7 protein is a fascinating approach.

P7 has two Post-translational modification site prediction, Myristyl site at (20-25, 35-40, 42-47, 80-85 and (104-109) which allows weak protein-protein and protein-lipid interactions (Farazi, 2001) and plays an essential role in membrane targeting, protein-protein interactions and functions widely in a variety of signal transductionpathways.Camp_phospho_site at (115-118) is implicated in the regulation of smooth muscle relaxation and activation of p7 channels for assembly and release of infectious virions. Expression of p7 channels was sufficient to increase production of IL-1β (Burdette *et al.*, 2012), likely through activation of NOD-like receptors (NLRP3) inflammasomes in primed macrophages. Our results suggest that active p7 channels are also expressed on the outer cell membrane and contribute to liver inflammation (Farag *et al.*, 2013).

Conclusion

The domain prediction is the first step for target structure prediction. In this work, Different bioinformatics servers were used to predict and analyze the p7-transactivated protein 1 and to discover protein binding motifs relating to its biological functions. Also, this study included the prediction of family conserved regions, secondary structure, solvent accessibility, tertiary structure, interaction motifs, post-translational modification sites and

disordered regions. Only one domain (1-127) was predicted and a fairly best model had been obtained from I-TASSER server after refinement and energy minimization according to best value of C-score, RMSD, TM-score and QMEAN Z-score. P7 has two Post-translational modification site prediction: Camp_phospho_site and Myristyl site, which has ion channel/pore-like function and crucial for production and release of infectious HCV particles from infected cells. Using SCOP, only one hit matched protein sequence at 71-120 and was classified as small proteins and FYVE/PHD zinc finger super family.

From CATH, sun protein; domain 3 super family has been used to demonstrate the functionality of the Ribosomal RNA small subunit methyltransferase B family. These results in the present study suggest that active p7 channels are also expressed on the outer cell membrane and contribute to liver inflammation.

ACKNOWLEDGEMENTS

The authors would like to express thanks to **Dr. Wessam H. El Behaidy** Lecturer, Computer Science Dept., Helwan University, Ain Helwan, Egypt, for help, valuable discussion, valuable advice, constructive effort.

REFERENCES

Atkinson, R.A., Williams, R.J., 1990. "Solution structure of the kringle 4 domain from human plasminogen by 1H nuclear magnetic resonance spectroscopy and distance geometry". *J. Mol. Biol*., 212 (3): 541–552.

Attwood, T.K., Bongcam-Rudloff, E., Brazas, M.E., Corpas, M., Gaudet, P., Lewitter, F., Mulder, N., Palagi, P.M., Schneider, M.V., van Gelder, C.W., GOBLET Consortium., 2015. "Correction: GOBLET: The Global Organisation for Bioinformatics Learning, Education and Training.". *PLoS Comput. Biol*., 11(5): e1004281.

Benkert, P., Biasini, M., Schwede, T., 2011. "Toward the estimation of the absolute quality of individual protein structure models". *Bioinformatics*, 27(3): 343-350.

Benkert, P., Tosatto, S.C.E., Schomburg, D., 2008. "QMEAN: A comprehensive scoring function for model quality assessment." *Proteins: Structure, Function, and Bioinformatics*, 71(1): 261-277.

Berjanskii, M., Liang, Y., Zhou, J., Tang, P., Stothard, P., Zhou, Y., Cruz, J., Macdonell, C., Lin, G., Lu, P., Wishart, D.S., 2010. *PROSESS*: A protein structure evaluation suite and server, Nucleic Acids Res. Web server Edition;

Bhattacharya, D., Cheng, V., 2012. 3-Drefine: Consistent Protein Structure Refinement by Optimizing Hydrogen-Bonding Network and Atomic-Level Energy Minimization. *Proteins: Structure, Function and Bioinformatics*. (In Press).

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L., Schwede, T., 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.*, 42: 252-258.

Burdette, D., Haskett, A., Presser, L., McRae, S., Iqbal, J., Waris, G., 2012. Hepatitis C virus activates interleukin-1beta via caspase-1-inflammasome complex. *J. of General Virology*, 93(2):235-46

Catherine, B., Emmanuel, C., Sébastien, C., Yoann, B., Sandrine, D., Daniel, K., 2005. The ProDom database of protein domain families: more emphasis on 3D. *Nucleic Acids Res.*, 33: 212-215.

Cheng, J., Baldi, P., 2007. Improved residue contact prediction using support vector machines and a large feature set. *BMC Bioinformatics*, 8:113.

Cheng, V., Randall, A., Sweredoski, M., Baldi, P., 2005. Protein Structure and Structural Feature Prediction Server, *Nucleic Acids Res.*, 33: 72-76.

Combet, C., Blanchet, C., Geourjon, C., Deléage G., 2000. NPS@: Network Protein Sequence Analysis. *Trends Biochem. Sci.*, 25 (3): 147-50.

Drozdetskiy, A., Cole, C., Procter, J., Barton, G.J., 2015. JPred4: A protein secondary structure prediction server, *Nucleic Acids Res.*, 1.

Farag, N.F., Breitinger, U., Breitinger, H.G., El Azizi, M.A., 2013. The p7 viroporin of the hepatitis C virus contributes to liver inflammation by stimulating the production of Interleukin-1β. https:// www. escmid. org/escmid_library/online_lecture_library/material/?mid=26934 .

Farazi, T.A., 2001. "The Biology and Enzymology of Protein N Myristoylation". J. of Bio. Chem. 276 (43): 39501–39504.

Ginalski, K., Elofsson, A., Fischer, D., Rychlewski, L., 2003. "3D-Jury: a simple approach to improve protein structure predictions". *Bioinformatics*, 19 (8): 1015–1018.

Gower, E., Estes, C., Blach, S., Razavi-Shearer, K., Razavi, H., 2014. Global epidemiology and genotype distribution of the hepatitis C virus infection. اكمل

Hepatol, J. 61(1): 45–57. Graham, C.S., Swan, T., 2015. A path to eradication of hepatitis C in low- and middle income countries. Antiviral Res. 119: 89–96.

Hall, H., 2001. BioEdit Version 5.0.6. North Carolina State University, Department of Microbiology, اكمل

Henikoff, J.G., Greene, E.A., Pietrokovski, S., Henikoff, S., 2000. Increased coverage of protein families with the Blocks Database servers. *Nucleic Acids Res.*, 28, 228–230.

Henikoff, S., Henikoff, J.G. Pietrokovski, S., 1999. Blocks +: A non-redundant database of protein alignment blocks derived from multiple compilations. *Bioinformatics*, 15: 471–479.

Hooft, R.W., Vriend, G., Sander, C., Abola, E.E., 1996. "Errors in protein structures*." Nature* m 381(6580): 272.

Kahn, D., Rezvoy, C., Vivien, F., 2008. Parallel large scale inference of protein domain families. *Proceedings of the 14th International Conference on Parallel and Distributed Systems, Melbourne, Australia, IEEE*, 72-79.

Katoh, K., Kuma, K., Toh, H., Miyata, T. 2005. MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucl. Acids Res.*, 33: 511–518.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10, 845-858.

Klepeis, J.L., Wei, Y., Hecht, M.H., Floudas, C.A., 2005. Ab initio prediction of the three-dimensional structure of a de novo designed protein: a double-blind case study. *Proteins*, 58 (3): 560–570.

Ko, J., Park, H., Heo, L., Seok, C., 2012. GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Res.*, 40: 294-297.

Kumar, S., Stecher, G., Tamura, K., 2016. Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.*, 33(7):1870–1874.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. 'PROCHECK: A program to check the stereo chemical quality of protein structures', *J. Appl. Cryst.*, 26: 283-291.

Letunic, I., Doerks T., Bork, P., 2015. SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res*., 43: 257-260.

Lin, K., Simossis, V.A., Taylor, W.R., Heringa, J., 2005. A Simple and Fast Secondary Structure Prediction Algorithm using Hidden Neural Networks. *Bioinformatics,* 21(2):152-9.

Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J., Russell, R.B., 2003. Protein disorder prediction: implications for structural proteomics. *Structure*, 11 (11): 1453-1459.

Liwo, A., Khalili, M., Scheraga, H.A., 2005. Ab initio simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains. *Proc Natl. Acad. Sci.*, USA 102(7): 2362–2367

McGuffin, L.J., Bryson, V., Jones, V., 2000. The PSIPRED protein structure prediction server. *Bioinformatics*, 16: 404–405.

Montserret, R., Saint, N., Vanbelle, C., Salvay, A.G., Simorre J.P., Ebel, C., Sapay, N., Renisio, J.G., Böckmann, A., Steinmann, E., Pietschmann, T., Dubuisson, J., Chipot, C., Penin, F., 2010. NMR structure and ion channel activity of the p7 protein from hepatitis C virus. *J. Biol. Chem.* , 285: 31446–31461.

Moradpour, D., Penin, F., 2013. Hepatitis C virus proteins: From structure to function. *Curr. Top. Microbiol. Immunol.* ,369: 113–142.

Niepmann, M., 2013. Hepatitis C virus RNA translation. *Curr. Top. Microbiol. Immunol.*, 369: 143–166.

Pei, J., Grishin, N.V., 2007. "PROMALS: towards accurate multiple sequence alignments of distantly related proteins". *Bioinformatics*, 23(7):802-808.

Pollastri, G., Martin, A.J.M., Mooney, C., Vullo, A., 2007. "Accurate prediction of protein secondary structure and solvent accessibility by consensus combiners of sequence and structure information", *BMC Bioinformatics*, 8:201.

Pollastri, G., McLysaght, V., 2005. Porter: A new, accurate server for protein secondary structure prediction. Bioinformatics 21: 19–20.

Rost, B., Yachdav, G., Liu, J., 2004. The PredictProtein Server. Nucleic Acids Res. 32:321-326.

Sievers F., Wilm A., Dineen D., Gibson TJ., Karplus K., Li W., Lopez R., McWilliam H., Remmert M., Söding J., Thompson JD., Higgins DG. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol* .,7:539.

Sigrist, C.J.A., de Castro, E., Cerutti, L., Cuche, B.A., Hulo, N., Bridge, A., Bougueleret, L., Xenarios, I., 2013. New and continuing developments at PROSITE. *Nucleic Acids Res*, 41(Database issue): 344-347.

Simmonds, P., 2013. The origin of hepatitis C virus. Curr. Top. *Microbiol. Immunol*., 369, 1-15.

Sitao, W., Yang, Z., 2007. LOMETS: A local meta-threading-server for protein structure prediction. Nucleic Acids Res. 35 (10) 3375–3382.

Sugiyama, K., 2004. Genomic structure and function of untranslated region, structural region and nonstructural region of hepatitis C virus RNA. Nippon Rinsho. 62 (7): 32–37.

Wallner, B., Elofsson, A., 2005. Pcons5: combining consensus, structural evaluation and fold recognition scores. *Bioinformatics*, 21(23): 4248-4254.

Wan, J., Kang, S., Tang, C., Yan, J., Ren, Y., Liu, J., Gao, X., Banerjee, A., Ellis, L.B.M., Li, T., 2008. "Meta-prediction of phosphorylation sites with weighted voting and restricted grid search parameter selection", *Nucleic Acids Res*., 36: e22.

Wu, S., Zhang, Y., 2007. LOMETS: A local meta-threading-server for protein structure prediction. *Nucleic Acids Res*., 35: 3375-3382.

Wu, S., Zhang, Y., 2008. A comprehensive assessment of sequence-based and template-based methods for protein contact prediction. *Bioinformatics*, 24:924–931.

Wu, S., Zhang, Y., 2008. MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information. *Proteins: Structure, Function, and Bioinformatics*, 72: 547-556.

Xu, D., Zhang, D., 2012. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins*, 80, 1715-1735.

Xu, D., Zhang, Y., 2011. Improving the Physical Realism and Structural Accuracy of Protein Models by a Two-step Atomic-level Energy Minimization. *Biophysical Journal*, 101: 2525-2534.

Xue, Z., Xu, D., Wang, Y., Zhang, Y., Threa Dom. 2013. Extracting Protein Domain Boundary Information from Multiple Threading Alignments. *Bioinformatics*, 29: 247-256.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y., 2015. The I-TASSER Suite: Protein structure and function prediction. *Nature Methods*, 12: 7-8.

Zhang, Y., Skolnick J., 2004. Scoring function for automated assessment of protein structure template quality. *Proteins*, 1; 57(4): 702-10.

Zhang, Y., Skolnick, J., 2005. TM-align: A protein structure alignment algorithm based on TM-score, Nucleic Acids Research, 33: 2302-2309.

التنبؤ وتحليل تركيب الثلاثي الأبعاد للبروتين بي سبعة لفيروس الالتهاب الكبدي الوبائي

محمود الحفناوي، محمد حسن، آمال محمود، يحيي خضر، السيد العبساوي، علاء حميدة

معهد الهندسة الوراثية – جامعة المنوفيه – مدينة السادات.

تحليل أى بروتين مثل فيروس الكبد الوبائى سى، وعمل نموذج ثلاثى الابعاد لشكله الهيكلى ، يساعد فى فهم آليات التفاعل الخاصه به ويمدنا بوسائل لإجراء اختبارات لتحضير الدواء المناسب .ووفقالمنظمة الصحة العالمية أكثرمن 185مليون شخصمصابمزمنمعفيروس (سي) في جميع أنحاءالعالم، وقدر 350 ألف الي 500 لف حالة وفاة سنويابسببالإزمان، والتليف الكبد يوتليف الكبد،وعلى نحو متزايد، وسرطان الخلايا الكبدية.

هذا العمل يقوم باقتراح منهجية تكاملية تتكون من مجموعة متداخلة من البرام ج المعلوماتية الحيوية .تُمكن هذه المنهجية الحصول على العديد من المعلومات لإستخلاص بعض المعارف عن الوظائف أو الاشكال الهيكلية للبروتينات .وقد تم تطبيق هذه المنهجية لعمل نموذج للبروتين "بي7بنجاح وساهم فى فهم آليات التفاعل، وإيجاد علاقة بين السلاسل الأمينية بالسمات الهيكلية والهيكل الثلاثى الابعاد للعديد من الوظائف الخاصة بهذا البروتين.

التوصية: