J.Agric.Chem.and Biotechn., Mansoura Univ.Vol. 8 (9): 225 – 230, 2017

Phylogenetic Analysis and Secondary and Tertiary (3D) Structure Prediction of the p7 Protein of Hepatitis C Virus

Iman Mohamed[1] ; A. A. El-Morsi[1] ; H. E. Soweha[1] ; H. Zaghloul[2] and M. M. El Hefnawi[3]

[1] Botany Department, Faculty of Science, Mansoura University, Egypt

[2] Clinical Pathology Department, Faculty of Medicine, Mansoura University, Egypt

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

ABSTRACT

The p7 protein of Hepatitis C virus (HCV) is a small integral membrane protein consists of 63 amino acids that is crucial for assembly and release of infectious virions. The p7 protein forms viral ion channels that can change membrane permeability, and could thus be considered as an antiviral target for drug design.The present study was designed to carry out a phylogenetic analysis and secondary and 3-D structure prediction of the p7 protein in the HCV genotype 4a detected in different countries. This is to examine the diversity of this genotype in different geographical regions and determine conserved regions that can be considered as a potential antiviral target for drug design. Molecular evolutionary and phylogenetic analysis using Jalview program showed that the HCV p7 gene of the genotype 4a isolates detected in Egypt was closely related to their counterpart in Germany .The phylogenetic analysis of HCV p7 genotype isolates from several parts of the world showing high genomic diversity of genotype 4 where the HCV-p7 protein genotype 4a dendrogram illustrates that the Egyptian isolates were classified into four clusters .Secondary structure predictions using JNetpred algorithm suggested that the p7 protein of the HCV genotype 4 contain one alpha Helix at 20-34(HPRLVRHLLHLHC amino acids) and three β-sheets at 5-7(GSVamino acids), 13-16(QCCF amino acids) and 43-47(CCYLR amino acids. The 3-D structure prediction model for the p7 protein of different HCV genotype 4a isolates (using J.mol) showed two coiled-coils . The α-helical coiled coil is a principal subunit oligomerization motifs in approximately 10% of all protein sequences.

Keywords: Hepatitis C virus, p7 protein, protein structure prediction, phylogenetic analysis.

INTRODUCTION

Hepatitis C virus (HCV) infection is a serious health problem causing the chronically liver disease which is the main reason for morbidity and death (Cook *et al*., 2013). It was reported by the World Health Organization (WHO) that approximately4 million people are newly infected per year and more than 185 million people are chronically infected with HCV in the world with approximately(0.35-0.5) million deaths estimated annually(Gower *et al* .,2014;Graham and Swan, 2015.) This chronic liver disease is the leading cause of hepatic fibrosis, cirrhosis, liver failure, hepatocellular carcinom (HCC) and finally death (Sugiyama, 2004 ; Lauer and walker, 2001)

HCV virus is a member of the genus *Hepacivirus* in the *Flaviviridae* family (Simmonds *et al*., 2017).The genome of HCV is a 9.6 kb single strand RNA encoding a single polyprotein precursor (Simmonds , 2013) which is processed by both of host and viral proteases into 10 mature proteins (Moradpour and Penin, 2007; Niepmann, 2013) including three putative structural proteins (the Core protein, which composes the nucleocapsid; the viral envelope E1 and E2 glycoproteins), the p7 ion channel viroporin required for virus assembly and the replication machinery consisting of 6 non-structural proteins (NS2, NS3, NS4A, NS4B, NS5A, and NS5B) (Lohmann , 2013).

HCV p7 protein is a new antiviral target that was found to be a member of viroporins (Pavlovic *et al*., 2003). As an ion channel molecule, the p7 protein are able to change the permeability of cell membranes (Nieva *et al*., 2012;Scott and Griffin , 2015) and has sensitivity to a variety of inhibitors in vitro (Premkumar *et al*., 2004). It is a monomer of 63 amino acids, which are hydrophobic and may contain endoplasmic reticulum (ER) retention signals (Lin *et al*., 1994 and Haqshenas *et al*., 2007). The p7 protein is crucial for assembly and release of infectous viral particles (Montserret *et al.,* 2010; Gentzsch *et al*., 2013and Grower *et al*., 2014).

One of protein structure prediction software applications is the prediction of three dimensional (3-D) structures of virus proteins, hence elucidating their functions. This will aid in drug design, and integrative understanding of viral processes.

In this study, bioinformatics *in silico* systematic analysis of the HCV p7 protein was used to predict the secondary and the 3-D structure to provide better understanding of the structure of p7 protein. This has been achieved using a set of bioinformatics tools such as Jalview, J.Pred and Jmol.

METHODS

The p7 protein analysis including the prediction of conserved regions, secondary structures, three-dimensional structures were done. The sequences for p7 ion channel, each consisting of 63 amino acids, were retrieved from sequence database on the HCV database (https://hcv.lanl. gov/).

1. Multiple sequence alignment

Multiple sequence alignments of the p7 protein of the genotype 4a were conducted using Jalview 2.10.2 (www.jalview.org/) program (Waterhouse *etal*., 2009). in which Sequences can be aligned using a range of algorithms with JABA web services such as ClustalW, Muscle, ProbCons, MAFFT, T-COFFEE and Clustal Omega.The Clustal Omega server (Sievers *et al*., 2011) is the fastest and most accurate tool for protein multiple alignment. Clustal Omega is a multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to make alignments among three or more sequences. While the most widely used one, Clustal W (Thompson *et al*., 1994) was used

for multiple sequence alignment of the DNA nucleotide sequences of the p7 gene of genotype 4a which were downloaded also from the HCV database.

2. Molecular evolutionary and phylogenetic analysis

The evolutionary history was inferred on both DNA level (nucleotide sequences) and protein level (amino acid sequences).The process of creating phylogenetic trees was done using neighbor joining method on the Jalview 2.10.2 program (Waterhouse *etal*., 2009).The neighbor joining is a bottom-up clustering method for reconstructing the phylogenetic trees from evolutionary distance data, created by Saitou and Nei in 1987. Usually used for trees based on DNA or protein sequence data, the algorithm requires knowledge of the distance between each pair of taxa (*e.g*., species or sequences) to form the tree (Xavier *et al*., 2010).

3. Secondary structure and solvent accessibility prediction

For the most accurate prediction, the consensus sequence of all 27 downloaded p7 protein sequences of the genotype 4a from different countries including Egypt were used. The secondary structure prediction was obtained using Jpred4 server (Alexey *etal*., 2015). JPred4 (http://www.compbio.dundee.ac.uk/jpred4) is the newest version of the popular JPred protein secondary structure prediction server which provides predictions by the JN *et al* gorithm, one of the most accurate methods for secondary structure prediction , solvent accessibility predictions and coiled- coil regions. Also, the protein prediction 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).

4. 3D Structure Prediction

3D structure prediction of p7 protein was done using Jmol server (Herraez A, 2006) which is an open-source Java viewer for chemical structures in 3-D(http://www.jmol.org/). The consensus sequence of all 27 downloaded p7 protein sequences of the genotype 4a from different countries including Egypt was used.

RESULTS AND DISCUSSION

1. Multiple sequence alignment

Multiple sequence alignment was done on the FASTA format of p7 ion channel sequences for genotype [4], subtype [a] from different countries to demonstrate the effect of geoghraphical distribution on the genotype genetic diversity. It was done using the Jalview server (Waterhouse *etal*., 2009) for DNA as shown in figure 1 and for protein as shown in figure 2 .

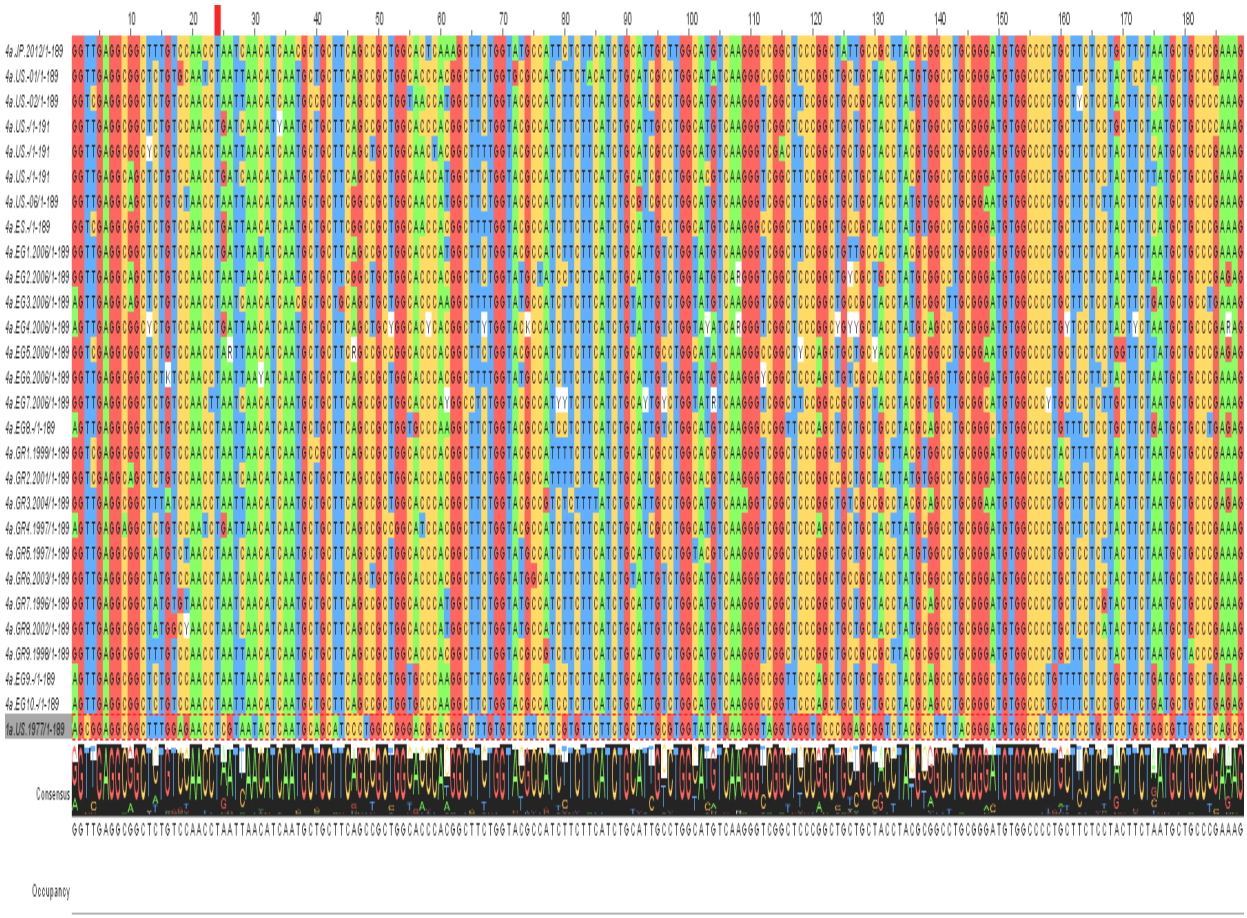

Fig. 1. Multiple sequence alignment of the nucleotide sequences of HCV- p7 gene for genotype [4], subtype [a [showing consensus sequence of p7 gene.

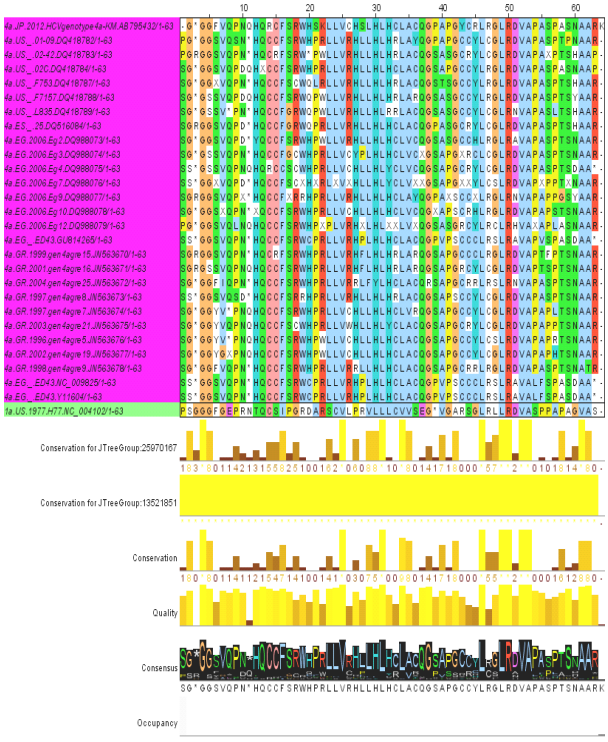

Fig. 2. Multiple sequence alignment of HCV-p7 protein (amino acid sequences) for subtype 4a showing conserved regions of p7 protein. Highly conserved regions were showed in yellow colored columns in the chart below the alignment.

Multiple sequence alignment of protein FASTA sequence for the complete sequence of p7 from genotype 4 (GT4), subtype [a] was also analyzed to illustrate the conserved regions. Mathew *et al*., 2015 showed less residues variation at the C-terminal region of p7 as compared to the loop region and N-terminal region in GT4 sequences. Comparative sequence alignment was also performed for GT4 p7 subtypes to identify the sequence with highly conserved residues which are the target regions for interaction with ligands as antiviral molecules. Therefore, subtype 4a (isolateED43) from genotype 4 have showed maximum similarity with other subtypes of the same genotype (Mathew *et al*., 2015).

2. Molecular evolutionary and phylogenetic analysis

The Jalview program was used for construction of phylogenetic tree of all HCV p7 sequences of the genotype [4a] from different countries to demonstrate the evolutionary relationship among them. The phylogenetic tree was constructed using Neighbor-Joining (NJ) method (Saitou and Neim, 1987). Phylogenetic analysis of p7 was done for both DNA as shown in figure 3 and for protein as shown in figure 4 .It revealed that HCV p7 protein genotype 4 of Egypt was closely related to p7 protein genotype 4 of Germany (4a.GR.2003).

This dendogram illustrates that the Egyptian isolates were classified into four clusters namely A, B, C and D. Cluster A containing Eg3, Eg4, Eg7 and it was found that Eg4 and GR 2003(Germany) making a clade of the same ancestor with different genetic distances (57.40 , 24.60) although Eg3 which have same genetic distance of Eg4 (57.40) but emerging from different ancestor and Eg7 is the most different one which has a separate branch emerging from another ancestor sharing with all of them . Cluster B containing Eg2, Eg10 that were very different and each one emerging in a separate branch with different genetic distance (48.47, 57.93). Cluster C containing (NC_009825, Y11604) more closely in one clade with same ancestor and same genetic distance (13.00) but both of them sharing a same ancestor with (GU814265) that has different genetic distance (14.44). Cluster D containing Eg9, Eg12 emerging from two separate branches with different genetic distances (99.83, 77.19). The phylogenetic analysis of HCV p7 genotype isolates from different strains has showed high genomic diversity of genotype 4, 5, and 6, low diversity of genotype1, and moderate diversity in genotype 2 and 3 (Mathew *et al*., 2015).

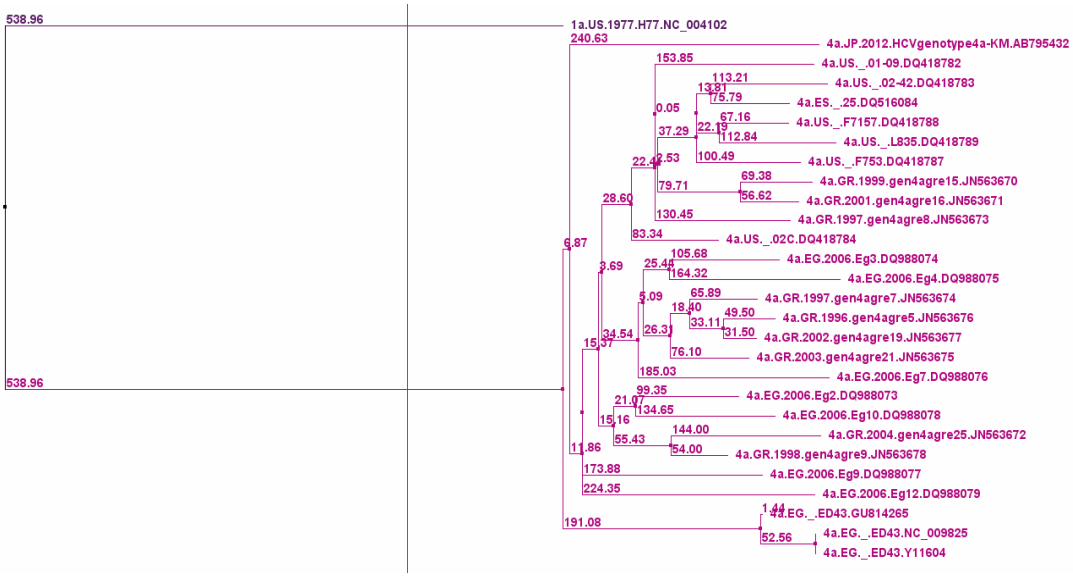

Fig. 3. Phylogenetic tree of HCV-p7 DNA genotype 4a from different countries using neighbor joining method .The numbers above the branches are the maximum likelihood branch support values that are obtained by using Jalview program.

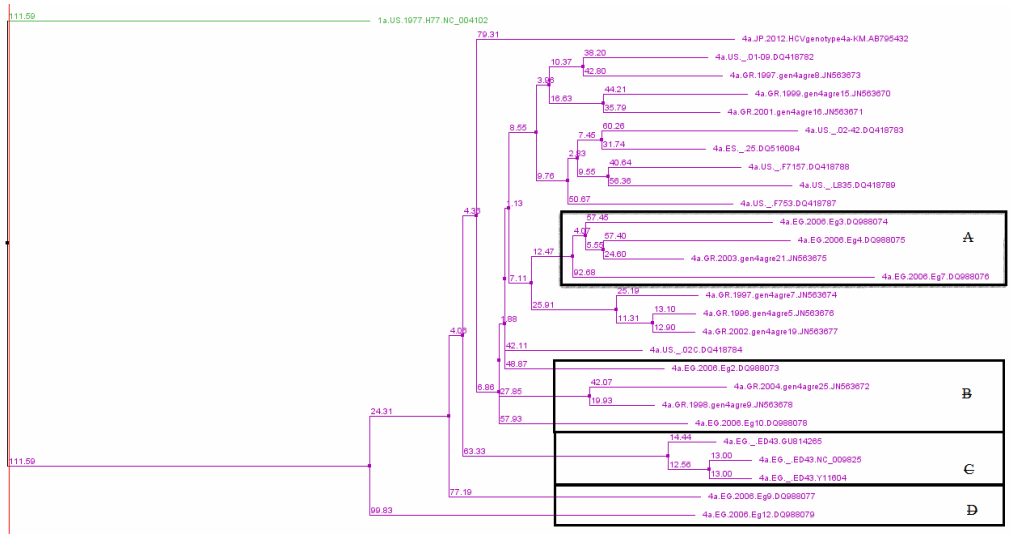

Fig. 4. Phylogenetic tree of HCV-p7 protein genotype 4a from different countries using neighbor joining method by BLOSUM62 .The numbers above each branch refer to the maximum likelihood branch support values that were obtained by using Jalview server.

For more illustration, Phylogenetic tree was done for the Egyptian isolates only as shown in figure5 which confirm the highly genetic diversity of HCV genotypes according to the geographical distribution.

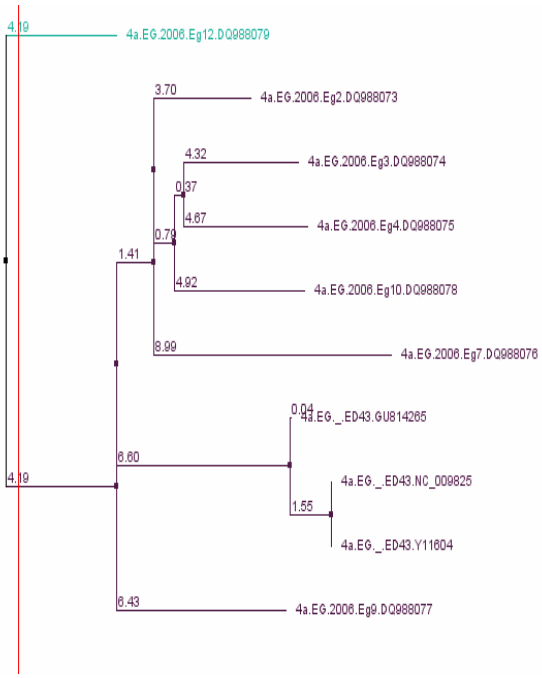

Fig. 5. Phylogenetic tree for only the HCV Egyptian isolates showing the genetic variations of p7 gene, genotype 4a from different isolates in Egypt.

3. Secondary structure and solvent accessibility prediction

To increase the accuracy of secondary structure prediction, we used the consensus sequence of all p7 protein sequences of the genotype 4a from different countries as input in the J.Pred 4 program for secondary structure prediction .JPred4 (http://www.compbio.dundee.ac.uk/jpred4) is a new version of the popular JPred protein secondary structure prediction server which provides predictions by the JN *et al* gorithm, one of the most accurate methods for protein secondary structure prediction.

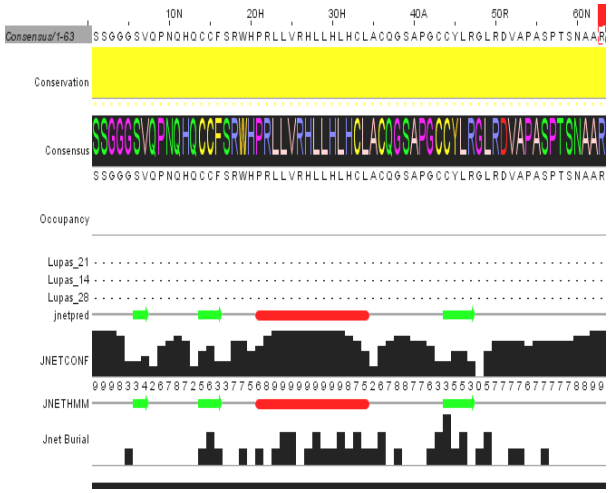

Fig. 6. The Secondary structure prediction for the consensus sequence of HCV p7 protein (consisting of 63 amino acids) genotype 4 , subtype a isolates from different countries in the world (using JPred4) revealed that it has one alpha helix and three beta sheets .Note : Beta sheets: green arrows. Alpha helices: red tubes.

HCV p7 protein for GT4 subtype (a) secondary structure prediction illustrate the presence of three Coiled-coil *i.e.*Lupas_21, Lupas_14, Lupas_28 as binary predictions for each location. JNetpred algorithm shows that Secondary structure of p7 protein composed of alpha helices ,beta strands and random coils.It has one alpha Helix that was specified at 20-34(HPRLVRHLLHLHC amino acids) and three *β*-sheets at 5-7(GSVamino acids), 13-16 (QCCF amino acids) and 43-47(CCYLR amino acids) (figure 6) (Rost *et al*., 2004 and Drozdetskiy *et al*., 2015). The confidence estimation for the prediction using JNetCONF was found that the alpha helix region exhibits the highest confidence degree in this prediction. JNetHMM (HMM profile based prediction) confirm the evaluation of

this prediction. Finally the solvent accessibility levels (S.A.) of the predicted protein interaction motifs of the HCV-p7 protein using Jnet Burial were found to be mostly exposed 66%, intermediate 25% and buried 9% (El Hefnawi *et al*., 2017).

4. 3D Structure Prediction:

The 3D structure prediction of the consensus sequence of all p7 protein genotype 4a from different countries was done using Jmol server (Herraez A, 2006) where several initial models are set up ,refined, evaluated and then the highest quality one is selected and the resulting structure was shown in figure 6.

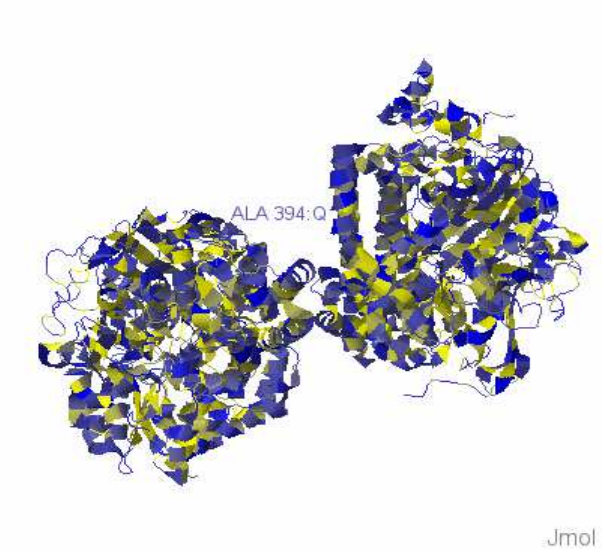

Fig. 7. 3-D (tertiary) structure prediction model for the p7 protein of different HCV- 4a isolates (using J.mol) showing two coiled-coils .

3-D structure validation of genotyoes 4 and 3 was done using 3-D structure template (Maiti *et al*., 2004).The α-helical coiled coil is a principal subunit oligomerization motifs in approximately 10% of all protein sequences. It is a highly versatile folding motif formed by 3 - 5 % of all the amino acids in proteins (Wolf *et al*., 1997). Coiled-coil proteins is important for different functions depending on the specific 'design' of their coiled-coil domains (e.g. gene regulation, cell division, membrane fusion and drug extrusion) and this demonstrate the importance of highly specific oligomerization in all biological systems (McFarlane *et al*., 2009). The classical coiled-coil domain includes a series of heptad repeats in the protein sequence that consists of two to seven alpha-helices wrapped around each other into a left-handed supercoil helix (Mason and Arndt, 2004) that are readily identifiable by the location of hydrophobic residues at the 'a' and 'd' positions (Burkhard *et al*.,2001). The heptad repeat occurs every two turns of the helix (Landschulz *et al*., 1988 and Lupas *et al*., 1996).

ACKNOWLEDGEMENTS

Authors honored to express their deepest gratitude and appreciation to Prof. Dr. Ahmed Mansour Alzohairy, Professor of genetics, and general manager of Bioinformatics Training Unit (BTU), Genetics Department, Faculty of Agriculture, Zagazig University for continuous assistance, useful discussions and revising this manuscript.

REFERENCES

Alexey, D. ; Christian, C. ; James, P. and Geoffrey, J. (2015). JPred4: a protein secondary structure prediction server .Nucleic Acids Research, 43: 389–394.

Burkhard, P.; Stetefeld, J. and Strelkov, S. V.(2001). Coiled coils: a highly versatile proteinfolding motif.TRENDS in Cell Biology, 11:82-88.

Clamp, M.; Cuff, J.; Searle,S. M. And Barton, G. J. (2004). The Jalview Java alignment editor". Bioinformatics. 20 (3): 426–7.

Cooke, G.; Lemoine, M.; Thursz, M.; Gore, C.; Swan, T.; Kamarulzaman, A.; DuCros, P. and Ford, N. (2013). Viral hepatitis and the Global Burden of Disease: a need to regroup. J. Viral Hepat., 20: 600-601.

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

El Hefnawi, M. M.; Hasan, M. E.; Mahmoud , A.; Khidr, Y. A.;El-Absawy, E. A.and Hemeida ,A. A. (2017). 3-D structure prediction and analysis of the p7-Transactivated protein 1of hepatitis C virus. J. Product. & Dev. 22(1): 83-102.

Gentzsch, J.;Brohm, C.;Steinmann, E.;Friesland, M.; Menzel, N.; Vieyres, G. ; Monteiro, P.;Perin ; Frentzen, A. ; Kaderali, L.and Pietschmann T. (2013). Hepatitis C Virus p7 is Critical for Capsid Assembly and Envelopment. PLoS Pathog.,9(5).

Gower, E.; Estes, C.;Blach, S.;Razavi-Shearer, K. and Razavi, H. (2014).Global epidemiology and genotype distribution ofthe hepatitis C virus infection. J.Hepatol. 61:45–57.

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

Haqshenas, G.; Mackenzie, J. M.; Dong, X. and Gowans E. J. (2007). Hepatitis C virus p7 protein is localized in the endoplasmic reticulum when it is encoded by a replication-competent genome. J. Gen. Virol. 88:134-142.

Herraez, A. (2006). Biomolecules in the computer: Jmol to the rescue. Biochemistry and Molecular Biology Education, 34(4): 255-261.

Landschulz, W. H.; Johnson, P. F. and McKnight, S. L. (1988). Science., 240: 1759-1764.

Lauer, G. M. And Walker, B. D. (2001). Hepatitis C virus infection. N Engl J Med., 345:41-52.

Lin , C. ; Lindenbach, B. D.;Pragai,B.M.;McCourt,D.W.and Rice, C.M. (1994). Processing in the hepatitis C virus E2-NS2 region: identification of p7 and two distinct E2-specific products with different C termini, J. Virol. 68: 5063–5073.

Lohmann, V. (2013). Hepatitis C virus RNA replication. Curr. Top. Microbiol. Immunol., 369: 167–198.

Lupas, A. (1996).Coiled coils: new structures andnew functions .Trends Biochem. Sci., 21: 375 - 382.

Maiti, R.;Van Domselaar, G.H.; Zhang, H.,;Wishart, D. S.(2004). SuperPose a simple server for sophisticated structural superposition. Nucleic Acids Res. 32(Web Server issue)W590–4 .

Mathew ,S ;.Kaneez ,F ;.Fatmi ,M. Q ;.Archunan ,G ;.Ilyas , M ;.Begum,N ;.Azhar ,E ;.Damanhouri ,G. and Qadri ,I.(2015). Computational Docking Study of p7 Ion Channel from HCV Genotype 3 and Genotype 4 and Its Interaction with Natural Compounds. PLOS ONE ,1-26.

McFarlane,A. A., Orriss,G. L. and Stetefeld, J.(2009).The use of coiled-coil proteins in drug delivery systems. European Journal of Pharmacology , 625: 101-107.

Montserret, R.; Saint, N.; Vanbelle, C.; Salvay, A.; Simorre, J.; Ebel, C.;Sapay, N.; Renisio, J.;Böckmann , A.; Steinmann, E.; Pietschmann, T.;Dubuisson, J.; Chipot, C. and Penin, F. (2010). NMR structure and ionchannel activity of the p7 protein from hepatitisCvirus. J. Biol. Chem. 285: 31446-31461.

Moradpour, D.; Penin, F. and Rice, C. (2007). Replication of hepatitis C virus. Nat. Rev. Microbiol., 5: 453–463.

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

Nieva, J.; Madan, V. and Carrasco, L. (2012). Viroporins: Structure and biological functions. Nat. Rev. Microbiol. 10: 563–574.

Pavlović, D.; Neville, D.C.; Argaud, O.; Blmberg, B.; Dwek, R.; Fischer, W. and Zitzmann, N.(2003).The hepatitis C virus p7 protein forms an ion channel that is inhibitedby long-alkyl-chain iminosugar derivatives, Proc. Natl. Acad. Sci. U. S. A. 100: 6104–6108.

Premkumar, A.; Wilson, L.; Ewart, G.D. and Gage, P.W. (2004).Cation-selective ion channelsformed by p7 of hepatitis C virus are blocked by hexamethyleneamiloride, FEBSLett. 557: 99-103.

Robertson, B.; Myers, G.; Howard, C. (1998). Classification, nomenclature, and database development for hepatitis C virus (HCV) and related viruses: proposals for standardization. International Committee on Virus Taxonomy. Arch Virol, 143: 2493 –2503.

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

Saitou, N. and Neim M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution, 4(4): 406-425.

Scott, C. and Griffin, S. (2015).Viroporins: Structure, function and potential as antiviral targets. Journal of General Virology. 96: 2000–2027.

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

Simmonds, P.(2013).The origin of hepatitis C virus. Curr. Top. Microbiol. Immunol., 369: 1–15.

Simmonds, P.; Becher, B.; Bukh, J.; Gould, E.A.; Meyers, G.; Monath, T.; Muerhoff, S.; Pletnev, A.; Rico-Hesse, R.; Smith, D.B.; Stapleton, J.T. and ICTV Report Consortium. (2017). ICTV Virus Taxonomy Profile: *Flaviviridae*. Journal of General Virology, 98:2-3.

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.

Thompson, J. D.; Higgins, D.G. and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Research, 22:4673- 4680.

Waterhouse, A. M.; Procter, J. B.; Martin, D. M.; Clamp, M.; Barton, G. J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics. 25 (9): 1189–91.

Wolf, E.; Kim, P. and Berger, B. (1997). MultiCoil: a program for predicting two- and three-stranded coiled coils. Protein Sci., 6: 1179 – 1189.

Xavier, D. (2010). Sequence-Based Analysis of Bacterial Population Structures. In:Bacterial Population Genetics in Infectious Disease .Robinson, D.; Falush, D. and Edward, J. F. (eds.). John Wiley and Sons, 46–47.

تحليل النشوء والتنبؤ بالهياكل الثانوية وثلاثية الأبعاد (3D) للبروتين p7 من فيروس التهاب الكبد الوبائي C

إيمان محمد[1] ، عادل احمد على المرسى[1] ، هانى اميل سويحة [1] ، حسام زغلول[2] و محمود الحفناوى[3]

[1] قسم النبات ، كلية العلوم ، جامعة المنصورة ، مصر.

[2] قسم الباثولوجيا الأكلينيكية، كلية الطب ، جامعة المنصورة ، مصر.

[3] قسم المعلوماتية الحيوية ، شعبة البحوث الهندسية ، المركز القومى للبحوث ، مصر.

يعد البروتين p7 لفيروس الالتهاب الكبدي سي أحدالبروتينات الغشائيةالمتداخلة صغيرة الحجم حيث انها تتكون من 63 حمض أميني وتلعب دور رئيسي في تجميع وتحرير الدقائق الفيروسية النشطة. هذا ومن الثابت ان p7 بروتين يكون قناة أيونية تتحكم في نفاذية الغشاء، وبالتالي p7 يعتبر أحد اهم المضادات الفيروسية الجديدة فضلا عن اهميته لتصميم الدواء المناسب. ويمكن تحقيق ذلك من خلال تحليل التطور الوراثي للنمط الجيني 4أ من مختلف البلدان وكذا فهم بنية البروتين p7 من خلال التنبؤ بالهيكل الثانوي والثلاثى الأبعاد له ومن ثم دراسة تأثير التوزيع الجغرافى على التنوع الوراثي و تحديد التتابعات الثابتةبها والتي ستكون هدف لتصميم العلاج الضد فيروسي .تحليلا النشوء والتطور الجزيئى للبروتين p7 أثبتا تقارب النمط الجينى 4 أ الموجود بمصر مع نظيره الموجود بألمانيا. كما اوضحت النتائج التنوع الوراثي الكبير بين العزلات الدولية وعليه تم تقسيم العزلات المصرية الي اربع محموعات. و بدراسة التركيب الثانوى للبروتين p7 للنمط الجينى 4أ وجد انه يحتوى علي حلزون ألفا واحد عند حمض اميني 20 الي 34 (HPRLVRHLLHLHC) وعلي ثلاثة من رقائق بيتا عند حمض اميني 5 إلي 7 (GSV)، وعند حمض أميني 13 إلي 16 (QCCF) و عند حمض أميني من 43 إلي 47 (CCYLR). وأظهرت دراسة الهيكل الثلاثي الأبعاد للبروتينp7 لعزلات مختلفة من النمط الجينى 4أ أنه عبارة عن حلزونين ملتفين. ويعتبر حلزون ألفا هو الوحدة الأساسية للسمات الهيكلية ويشغل حوالي 10% من كل التتابعات البروتينية.