Helix (2022) 12 (1): 27-31

https://doi.org/10.29042/2022-12-1-27-31

In Silico Vaccine Designing Against Hepatitis C Virus

*[1]G Jyothsna, [2]Vinay Marappu
Bioaxis DNA Research Centre, Hyderabad, India
jyothsna@dnares.in
TSR and TBK Degree College, Visakhapatnam, A. P, India

Received: 17th December 2021, Accepted: 19th February 2022, Published: 28th February 2022

Abstract
Hepatitis is an infection of the liver caused by Hepatitis C virus. Though it's a viral infection affecting many individuals there is no good vaccine till date to prevent the infection. With the development in Science and technology, advancement in the protein databases and bio informatics application, designing a peptide vaccine has become more authentic and accurate for the prevention of these viral infections. Paper involves the proteomic analysis of the virus and identification of potential antigenic proteins and their pathogenic peptides in a series of rational steps. Once the antigenic peptides are identified they are filtered and selected based on their degree of antigenicity called as antigenic propensity. The final peptide selected with highest antigenic propensity is subjected for 3D structure designing in Argus lab software and the energy is calculated. This peptide can be further processed in laboratory to develop into a suitable peptide vaccine to prevent the infection.

Keywords: *Proteomics, Annotation, Hepatitis C, Pathogenic Peptides*

Introduction
Hepatitis is a viral liver infection causing swelling and malfunctioning of the liver. It is caused by a pathogenic virus Hepatitis C [1]. Some common forms of hepatitis include Hepatitis A, B and C. All the variants of the disease exhibit similar kind of symptoms but have different treatment protocol with different degree of pathogenecity [2].

The spread of infection from one person to another may occur due to contact with the infected blood. Hereditary from mother is one of the major causes of the infection [3]. Improper handling of the healthcare equipment and negligence may also be a cause of spread. It also exhibits sexual transmission from one partner to the other. The condition may remain asymptomatic in several cases [4]. However the most commonly observed symptoms include stomach disturbance, bloating, loss of appetite, yellowing of the skin and eyes, fever, sickness, dark urine color etc. Victims of chronic hepatitis C often go unnoticed due to lack of symptoms, however the symptoms can be experienced in these individual in most advanced cases with no hope for treatment.

Currently there is no vaccine to prevent the infection; however several measures can be adopted to prevent the spread of the infection from one to other. Some of the precaution to be adapted include: Avoid reuse of needles, syringes and any other medical equipment once used. It is also necessary to avoid sharing of anything with the infected person as the disease is contagious.

In silico Vaccine designing also termed as reverse vaccinology is a bioinformatics protocol involving the analysis of complete proteome of the pathogen [5]. Screening of the proteome helps in the identification of pathogenic proteins foreign to the human proteome. Based on these proteins a secondary screening is performed for the identification of antigenic peptides and their antigenic propensity calculation. Thus the protocol involves the identification of highly antigenic and immune provoking peptides whose synthesis and modification can yield a good peptide vaccine candidate for prevention of the infection.

Materials and Methods
Screening of Viral proteome: The complete set of proteins belonging to the virus are individually collected and filtered to select non repetitive, non hypothetical, complete proteins. These selected proteins can be

subjected for initial screening. Basic filtration criteria here are to discard all he proteins with the prefixes like, hypothetical, unnamed, predicted, partial etc. All the data is collected from NCBI [6] data base.

Identification of Foreign Proteins: For any protein to be pathogenic it is an essential criteria that the protein should be foreign to the host, thus can elicit and immune response. So it is an important step to identify the list of pathogen's proteins which are foreign to the host. This can be achieved using BLAST P [7] tool of NCBI server. This tool performs a pair wise sequence alignment followed by comparison of user entered protein sequences with the selected data base sequences. Here the data base to be screed must be selected as human proteome. All the viral proteins that either share no similarity or less than 30% similarity to the human proteome can be selected as the foreign protein. These foreign proteins can be subjected for further screening to identify most antigenic and potential vaccine candidates.

Antigenic peptide prediction using PVS and EMBOSS ANTIGENIC: Antigenecity of any protein is not the function of the whole unit but is limited to the regions on the protein called antigenic peptides. These antigenic peptides can be identified using PVS [8] server for the protein annotation. The tool not only provides an insight to the antigenic peptides but also displays a graph with the antigenic propensities of the identified peptides. The peaks in the graph are the indication of most antigenic regions within the protein. These results can be useful in the identification of final list of vaccine candidates.

EMBOSS ANTIGENIC [9] is another tool that helps in the identification of antigenic peptides along with the specific site within the region identified to be antigenic site based on Kolaskar and Tongaonkar method. Scores are also provided for each peptide to estimate its antigenic propensity. Both the results can be compared to select the final peptide list for vaccine designing.

3D Structure development of the final antigenic peptides as vaccine candidates: Once the final list of peptides is selected based on their antigenic propensity, they are subjected for 3D structure development using Argus lab software [10]. Argus lab is standalone software for modeling of proteins and chemical structures virtually. He structures thus developed can be used for various docking and interaction studies. Protein protein super impositions, energy calculations etc., can also be performed in argus lab.

Results and Discussion

After selecting all the proteins of the Virus excluding the repetitive and hypothetical proteins, these list of proteins were subjected for BLASTP with homo sapiens to identify the foreign proteins. He result shown below in Fig 1 is an example to the foreign protein showing no similarity to the human proteome. All such proteins of the virus can be selected for the study.

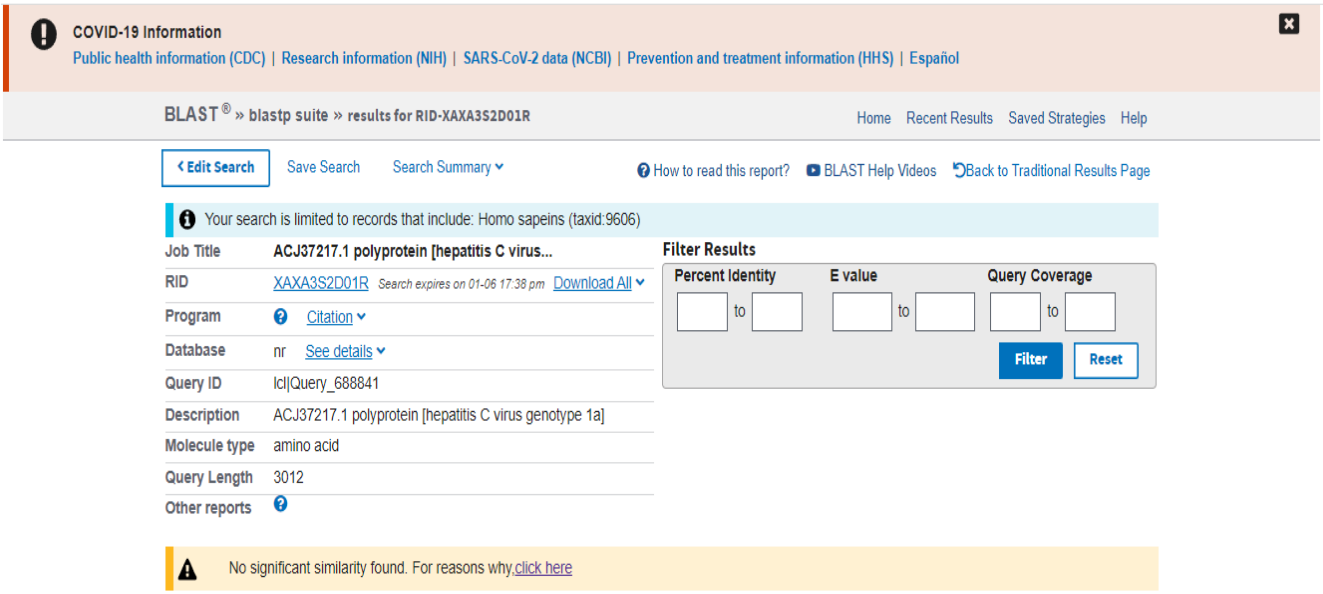

Fig 1: Protein showing no similarity to human proteome

Inference: The above figure 1 shows that the user entered protein is one of the proteins of the virus that shares no similarity to the human proteome. Thus it is foreign and can be selected for further screening.

All such proteins which share no similarity to the human proteome are to be selected for next screening. Those that share some degree of similarity must be discarded from the list in this step.

Polyprotein of Hepatitis C virus was found to be foreign and thus selected for further study. The sequence of polyprotein was subjected for PVS and EMBOSSES tools for the identification of antigenic peptides. Following results were obtained in both the analyses.

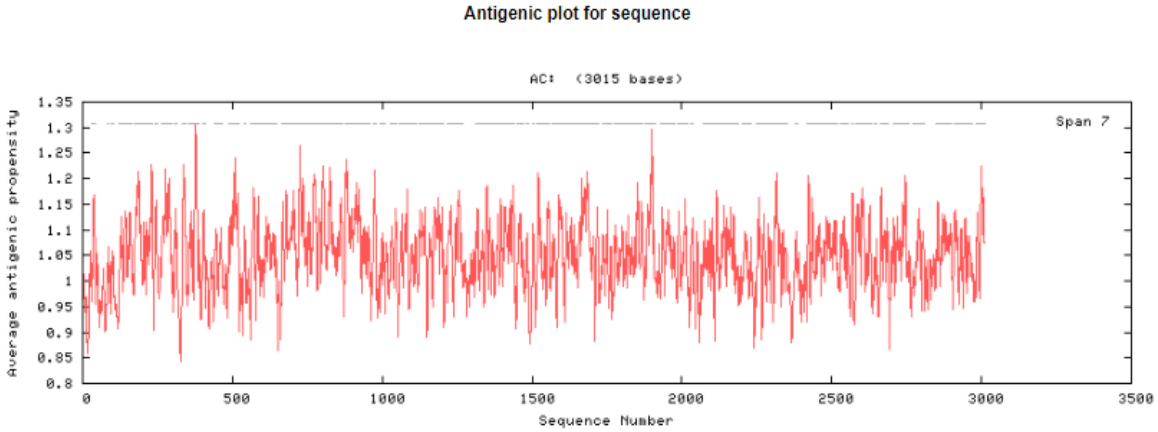

Fig 2: Plot of antigenic peptides against their propensity

Inference: From the above antigenic site plot it can be revealed that the region from 0 to 500 is touching the peak indicating its highest antigenic propensity among the other peptides. Further the region lies approximately from 250 to 500.

n	Start Position	Sequence	End Position
1	31	GGQIVGGVYLLP	42
2	82	QPGYPWP	88
3	97	WAGWLLS	103
4	122	NLGKVIDTLTCGF	134
5	136	DLMGYIPLVGAPLGGAARALAHGVRVLED	164
6	172	NLPGCSFSIFLLALLSCLTVPASAYQV	198
7	202	SGLYHVTN	209
8	212	PNSSIVYEAADAILHTPGCVPCVR	235
9	240	SRCWVAVTPTV	250
10	262	LRRHIDLLVGSATLCSALYVGDLCGSVFLVGQLFT	296
11	306	QDCNCSIYPGH	316
12	332	PTAALVVAQLLRIPQ	346
13	354	GAHWGVLAGIAYFS	367
14	371	NWAKVLVVLLLFAGVDAET	389
15	400	TAGLVGLLTP	409
16	441	WLAGLFYQH	449
17	452	NSSGCPERLASCRR	465
18	486	ERPYCWHYPPRPCGIVPAKSVCGPVYCFTPSPVVVG	521
19	538	TDVFVLN	544
20	563	GFTKVCGAPPCVIG	576
21	580	NNTLLCPTDCFR	591
22	597	TYSRCGS	603
23	607	ITPRCMVDYPYRLWHYPCTINYTIFKVRMYVGGVEHRLEA	646

Fig 3: PVS result showing the table of peptides along with their positions in the protein sequence

Inference: The above table details the position f each peptide that was shown in the previous graph. Now, by comparing the results from graph and table it can be clearly identified that the position 371- 389 is highest antigenic site with its propensity being nearly 1.3.

```
#      -minlen 5
#      -outfile outfile
#      -rformat2 motif
# Report_format: motif
# Report_file: outfile
#######################################

#---------------------------------------
#
# Sequence: NP_671491.1      from: 1    to: 3011
# HitCount: 127
#---------------------------------------

Max_score_pos at "*"

(1) Score 1.307 length 19 at residues 368->386
                    *
 Sequence: WAKVLVVLLLFAGVDAETH
             |                |
           368              386
 Max_score_pos: 374

(2) Score 1.295 length 33 at residues 1876->1908
                          *
 Sequence: TEDLVNLLPAILSPGALVVGVVCAAILRRHVGP
             |                              |
          1876                           1908
 Max_score_pos: 1895

(3) Score 1.265 length 31 at residues 716->746
                    *
 Sequence: WEYVVLLFLLLADARVCSCLWMMLLISQAEA
             |                            |
           716                          746
 Max_score_pos: 722

(4) Score 1.241 length 36 at residues 483->518
```

Fig 4: Identification of antigenic peptides and their scores by EMBOSS antigenic

Inference: In the result of EMBOSS antigenic the antigenic peptide identified was from 368 to 386 with a score of 1.307 and maximum antigenic score at the position 374.

SAA calculation of peptide NWAKVLVVLLLFAGVDAET

Total number of Valine + Cystine + leucine/ length of peptide = 4+ 4/ 19= 0.421

The Peptide sequence, NWAKVLVVLLLFAGVDAET of the protein Polyprotein [hepatitis c virus genotype 1a] is showing highest antigenic propensity in PVS and also high degree of SAA value when compared to the other peptides. Thus, it can be selected as final peptide vaccine candidate and thus can be subjected for 3D structure development in Argus lab.

3D structure development in argus lab: The 3D structure of peptide sequence was designed in Arguslab, and also the calculation of its energy was done. When a peptide shows the least energy possible, it would be considered most stable.

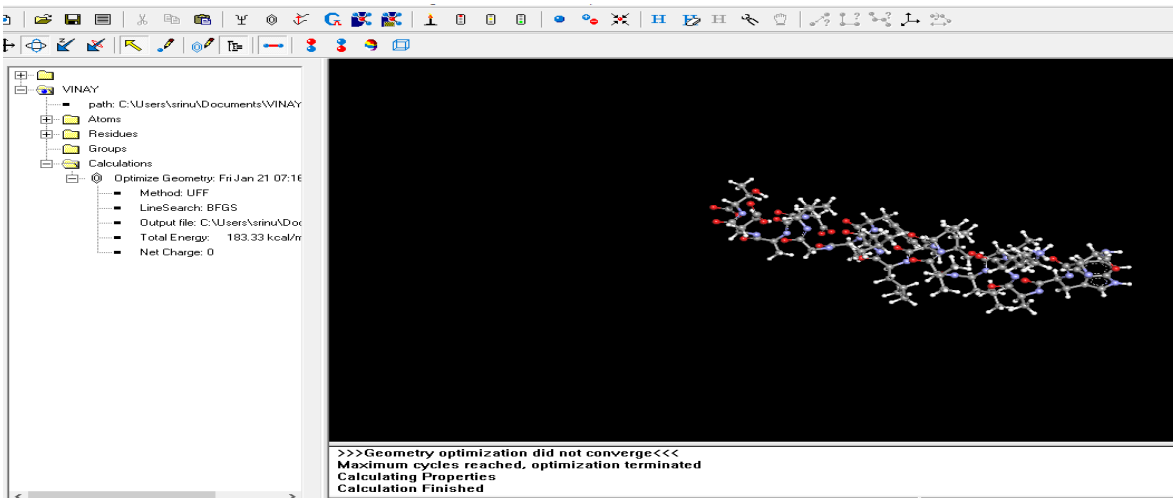

Fig 5: 3D structure of the selected vaccine candidate peptide

Conclusion

Hepatitis C Virus, of Flaviviridae family, is responsible for the cause of deadly liver diseases like cirrhosis, hepatocellular carcinoma, etc. This is a virus usually spread by direct contact with infected blood. Prevention is done by avoiding any kind of contact with infected blood and other ingesting materials.

In-silico vaccine designing is the process of synthesizing a peptide vaccine using bioinformatic tools. This process is cost effective and requires use of no harmful ways of synthesizing vaccine. Here, the vaccine is developed by using a peptide sequence which would be showing highest antigenicity against a human sequence. The vaccine developed by this method would be a poly peptide vaccine which shows no side effects upon ingestion.

In this research, a library of 20 protein sequences was collected from NCBI, which are from Hepatitis C Virus proteome. The proteins which are foreign to the human proteins were screened using BLAST-P tool. The screened proteins were run in PVS and Emboss to identify a peptide sequence with highest antigenicity and surface accessibility possible, this gives a peptide sequence which is foreign to human sequence, has antigenic epitopes and can be developed into a vaccine. Thus, it can be selected as final peptide vaccine candidate and can be subjected for 3D structure development in Argus lab. Hence, a peptide vaccine of sequence NWAKVLVVLLLFAGVDAET from the protein Polyprotein [hepatitis c virus genotype 1a], was developed with the least energy of 183.3 Kcal. However the sequence can be further synthesized in the laboratory and tested on lab animals to obtain a confirmatory result and use it for vaccination of individuals.

References

1. Kutscher S, Bauer T, Dembek C, Sprinzl M, Protzer U. Design of therapeutic vaccines: hepatitis B as an example. *Microb Biotechnol*. 2012;5(2):270-282. doi:10.1111/j.1751-7915.2011.00303.x

2. Irshad M, Mankotia DS, Irshad K. An insight into the diagnosis and pathogenesis of hepatitis C virus infection. *World J Gastroenterol*. 2013;19(44):7896-7909. doi:10.3748/wjg.v19.i44.7896

3. Elena Bevilacqua, Annalisa Fabris, Paolo Floreano, Lucy Pembrey, Marie-Louise Newell, Pier-Angelo Tovo, Antonio Amoroso, ”Genetic factors in mother-to-child transmission of HCV infection”, Virology,Volume 390, Issue 1, 2009, Pages 64-70, ISSN 0042-6822

4. William A. Sodeman Jr. M.D., J.D., F.A.C.P., F.A.C.G., F.A.C.L.M., Thomas C. Sodeman M.D., in Instructions for Geriatric Patients (Third Edition), 2005

5. Sette A, Rappuoli R. Reverse vaccinology: developing vaccines in the era of genomics. Immunity. 2010;33(4):530-541. doi:10.1016/j.immuni.2010.09.017

6. National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [cited 2017 Apr 06]

7. Zhang Z., Schwartz S., Wagner L., & Miller W. (2000), "A greedy algorithm for aligning DNA sequences" J Comput Biol 2000; 7(1-2):203-14

8. Garcia-Boronat M, Diez-Rivero CM, Reinherz EL, Reche PA. PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. Nucleic Acids Res. 2008;36(Web Server issue):W35-W41. doi:10.1093/nar/gkn211

9. Kolaskar AS, Tongaonkar PC (1990) A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Lett 276(1-2): 172–174

10. Thompson, M.A. (2004) Molecular Docking Using ArgusLab, an Efficient Shape-Based Search Algorithm and their Scoring Function. ACS Meeting, Philadelphia