Chapter 9
Identification of potential Vaccine Candidates for COVID19

The significance of vaccines in the battle of mankind with dreadful diseases cannot be surpassed. As it goes by saying ‘Prevention is better than cure’, vaccines have always proven their mettle. In view of the urge for the development of vaccine against the 2020 epidemic COVID the current work was undertaken. In continuation to the previous chapter on the identification of possible drug targets of COVID 19 [1] the current analysis aims to predict the possible antigenic peptides and sites on the ORF 1ab Polyprotein of Novel Corona Virus (Wuhan Isolate 2019). In view of its vital role in the replication of viral genetic material and its length covering a maximum of viral proteome, this protein is selected for the study. The protein sequence of ORF1ab poly protein was retrieved from NCBI and was subjected for BLAST analysis with Homo sapiens to identify its degree of foreignness to humans. Being a vital protein for replication it is expected to have high antigenicity and pathogenicity. It can be a better immunogen to trigger the production of antibodies by the host immune system. The study aims to develop potential vaccine candidates for stimulating the acquired immunity in the individual and prevent the adverse effects of the COVID infections.

As mentioned Corona Viruses are a group of RNA viruses [2] and pathogens of Mammals. They are further classified into various groups [3]. These are mostly winter viruses that show highest activity under low temperatures [4] however the temperature dependency is not complete. 2019–2020 can be marked as COVID era due to this major outbreak of COVID pandemic [5]. This pandemic had hit the entire world both socially and economically. Due to its contagious property it has currently become a major threat globally [6]. It has a slow activation period of 14–20 days before which the symptoms are not observed in the affected individuals. However these individuals can be good carriers for the virus even prior to its activation stage. Thus it is very difficult to control the spread of the pandemic.

Drugs can be developed to treat the infected individuals. However in order to protect the individuals from the virus vaccination is the only possible approach. Thus the current chapter involves the prediction of possible vaccine candidates/peptides against COVID using reverse vaccinology approach [7].
9.1 Determining the Foreignness of ORF1ab Poly Protein of Novel Corona Virus to Human Proteome Using BLAST

BLAST [8] is a basic local alignment search tool from NCBI. BLAST performs alignment for the comparison of any two sequences and identifies their degree of similarity. This tool is used to identify the degree of foreignness [9] between the human proteome and the targeted protein ORF1ab poly protein of Novel Corona Virus. In case the viral protein does not share any similarity, it would be recognized as FOREIGN to humans thus can be an immunogenic protein (Fig. 9.1).

**Inference:** The above BLAST result shows that the sequence of ORF1ab (Novel Corona Virus) shares 31% sequence similarity to Homo sapiens with a query coverage 1% only. This indicates that the sequence is not close to the human proteome and possesses foreignness; thus it can be antigenic and pathogenic and hence can be used for further analysis.

9.2 Identification of Antigenic Regions Using PVS

PVS [10] codes for Protein variability server that harbors many tools for the annotation of proteins. Antigenic Peptide prediction is a tool in PVS that would predict the antigenic regions within the protein having potential to trigger the production of antibodies in the host. Along with the antigenic peptides antigenic propensity of amino acids is also displayed in the tool output. The use of Peptide vaccines is always preferable than the whole organism Vaccine due to several parameters like less cost, high efficiency and lower risk of side effects (Fig. 9.2).
Fig. 9.2 Antigenic propensity plot of the sequence of ORF 1ab

The above graph of PVS server shows residues position number on the X axis and Antigenic propensity on Y axis. According to the graph the highest peak showing greater antigenic propensity is located between 5000 and 6000 position. This indicates that, target region to be used for vaccine peptide development lies in this region (Table 9.1).

PVS results revealed the presence of a total 311 antigenic peptides in the ORF 1ab poly protein. Among them the above table highlights the regions falling under the highest peak zone. The exact peptide having highest antigenic propensity is further identified using EMBOSS ANTIGENIC tool.

| Peptide Sequence | Position |
|------------------|----------|
| LAPAYVYF        | 5495     |
| YDIALVY        | 5539     |
| YGLVQGF        | 5583     |
| GYVYFCTVNA    | 5589     |
| TAOVTD        | 5701     |
| NYDLSWNA     | 5716     |
| LRAKYMPYCTFQAPRT | 5787   |
| FLGTCORCPAEVDTSYALVYD | 5785 |
9.3 Antigenic Site Prediction by EMBOSS ANTIGENIC

EMBOSS Antigenic [11] is used for the prediction of probable antigenic sites present in the above protein. The tool would predict the antigenic sites for the user entered protein sequence. The prediction is based on the Kolaskar and Tongaonkar’s method. This method of antigenic site prediction is based on the physiochemical properties of the amino acids. The output gives the information about antigenic region along with the indication of site within the region with high potential for being an antigenic site. All the data obtained was tabulated.

9.4 Identification of Peptide with Highest Antigenic Propensity Using EMBOSS Antigenic Tool

See Fig. 9.3.

The above EMBOSS results show that the peptide with highest score or antigenic propensity is the region 5337–5416. This shows the highest score of 1.284 confirming

```
# Program: antigenic
# Rundate: Tue 23 Jun 2020 11:54:42
# Commandline: antigenic
# -auto
# -sequence /var/lib/emboss-explorer/output/199995/.sequence
# -minlen 6
# -outfile outfile
# -rformat2 motif
# Report_format: motif
# Report_file: outfile

#******************************************************************************
# Seq: YP_009724389.1   from: 1 to: 7096
# HitCount: 323
#******************************************************************************

Max_score_pos at "*"
(1) Score 1.284 length 50 at residues 5337->5416
Sequence: SLLRGAICIRRPFLCCKCCYDMVISTHKLVLSDVNYFCNPVGCDVTVDQTLYLGGMSYCKSHKPPISFLCANGQVFGL
      5337      5416
Max_score_pos: 5352
(2) Score 1.276 length 14 at residues 4323->4336
Sequence: GASCCLYCCHIDH
      4323      4336
Max_score_pos: 4329
(3) Score 1.273 length 21 at residues 3864->3884
Sequence: OVKCTSVVLSVLQQLRVESS
      3864      3884
```

Fig. 9.3 Partial output of EMBOSS antigenic tool
it to be most immunogenic or antigenic peptide suitable for Vaccine development. However a total of 323 peptides were predicted to be antigenic using EMBOSS.

The results of EMBOSS and PVS indicate that the following peptide is the best target for the development of peptide based vaccine due to its high degree of antigenic propensity. The K in the pattern below is the site having maximum antigenic propensity or antigenicity as identified by EMBOSS tool (Fig. 9.4).

```
SLRCGACIRRFLCCKCYDHVISTSHKLVLSSNYVCNAPGCDVTVDVTQLYLGGMSYYCKSHKPPISFPLCANGQVF
```

**Fig. 9.4** Partial output of protparam tool for physicochemical characterization of peptide

| Amino acid composition: | CSV format |
|-------------------------|------------|
| Ala (A) 3 | 3.8% |
| Arg (R) 3 | 3.8% |
| Asn (N) 3 | 3.8% |
| Asp (D) 3 | 3.8% |
| Cys (C) 10 | 12.5% |
| Glu (Q) 2 | 2.5% |
| Glu (E) 0 | 0.0% |
| Gly (G) 6 | 7.5% |
| His (H) 3 | 3.8% |
| Ile (I) 3 | 3.8% |
| Leu (L) 8 | 10.0% |
| Lys (K) 4 | 5.0% |
| Met (M) 1 | 1.2% |
| Phe (F) 3 | 3.8% |
| Pro (P) 6 | 7.5% |
| Ser (S) 7 | 8.8% |
| Thr (T) 3 | 3.8% |
| Trp (W) 0 | 0.0% |
| Tyr (Y) 5 | 6.2% |
| Val (V) 7 | 8.8% |
| Pyl (O) 0 | 0.0% |
| Sec (U) 0 | 0.0% |

(B) 0 | 0.0%
(Z) 0 | 0.0%
(X) 0 | 0.0%

Total number of negatively charged residues (Asp + Glu): 3
Total number of positively charged residues (Arg + Lys): 7
The above results as obtained by protparam [13] annotate the peptide in detail. The length of the peptide is found to be 80 amino acids with the molecular weight of 8764.32 and isoelectric point 8.57. This is the pH at which the peptide becomes inactive in an electric field. The peptide was identified to be unstable with the Instability index being 47.53. The peptide is basic in nature.

9.5 Construction of Peptide 3D Structure in Argus Lab

The 3D structure of peptide predicted above was constructed in Argus lab. Argus lab [12] is downloadable software for protein and chemical annotation. It is specially used for targeted docking studies and molecule and protein building. It has an inbuilt programme of energy minimization and geometry optimization of the molecule. The energy of any molecule generated in the software can be calculated.

Figure 9.5 shows the 3D structure of the selected peptide which can be used as a potential vaccine candidate. The yellow highlighted site is amino acid K (Lys) which is the one with maximum antigenic propensity. The energy of the peptide was calculated to be 2523.94 kcal/mol.

![Fig. 9.5](image-url) The optimized structure of the vaccine candidate peptide designed in Argus lab
9.6 Conclusion

ORF1ab polyprotein of Novel corona virus (Wuhan Isolate 2019) was annotated further in the current research. Antigenic sites within the protein were predicted using EMBOSS ANTGENIC and were further validated by PVS (Protein variability Server). The regions showing variability are collected and subjected for further analysis. Results of the above tools were summarized to finalize a peptide with highest antigenic propensity. The peptide sequence was further annotated using Prot-param to understand the physicochemical properties of the protein. Argus Lab was used to build the 3D structure of the peptide and the energy was calculated to be 2523.94 kcal/mol. The study concludes that the selected peptide is a good vaccine candidate to activate the immune response in the healthy individuals.

The work is completely based on in silico analysis, hence further laboratory study and validation is necessary before testing the efficacy of the peptide as vaccine.

References

1. Gundlapally J, Kumar A, Kashyap A, Saxena AK, Sanyal A (2020) In search of novel coronavirus 19 therapeutic targets. Helix 10(02):01–08
2. Fehr AR, Perlman S (2015) Coronaviruses: an overview of their replication and pathogenesis. Methods Mol Biol 1282:1–23. https://doi.org/10.1007/978-1-4939-2438-7_1
3. Tyrrell DAJ, Myint SH (1996) Coronaviruses. In: Baron S (ed) Medical microbiology. 4th edn. University of Texas Medical Branch at Galveston, Galveston, TX. Chapter 60. Available from: https://www.ncbi.nlm.nih.gov/books/NBK7782/
4. Razuri H, Malecki M, Tinoco Y et al (2015) Human coronavirus-associated influenza-like illness in the community setting in Peru. Am J Trop Med Hyg 93(5):1038–1040. https://doi.org/10.4269/ajtmh.15-0271
5. Coronavirus: COVID-19 Is Now Officially a pandemic, WHO says, March 11, 202012:30 PM ET, The corona virus crisis. https://www.npr.org/sections/goatsandsoda/2020/03/11/814474930/coronavirus-covid-19-is-now-officially-a-pandemic-who-says
6. How Coronavirus Spreads, Coronavirus Disease 2019 (COVID-19), Centre for Disease Control and Prevention (CDC). https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covidspreads.html?CDC_AA_refVal=https%3A%2F%2Fhttps://www.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fprepare%2Ftransmission.html
7. Mora M et al (2006) Microbial genomes and vaccine design: refinements to the classical reversevaccinology approach. Curr Open Microbiol 9(5):532–536
8. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410. BLAST PROGRAMS
9. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 31:3784–3788
10. Garcia-Boronat M, Diez-Rivero CM, Reinherz EL, Reche PA (2008) PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. Nucleic Acids Res 36(Web Server issue):W35–W41. https://doi.org/10.1093/nar/gkn211
11. Rice P, Longden I, Bleasby A (2000) EMBOSS: The European molecular biology open software suite. Trends Genet 16(6):276–277
12. Book: Combined quantum mechanical and molecular mechanical methods (American Chemical Society ACS Symposium Series) 1997. ISBN-10: 0841235902
13. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. In: Walker JM (ed) The proteomics protocols handbook, Humana Press, pp 571–607