

CHAPTER 3
METHODOLOGY

3.1 Source of Flaviviruses sequences

A total of 5 sequences were used in the present study. The member of flaviviruses involved are Dengue Virus (DEV), West Niles Virus (WNV), Japanese Encephalitis Virus (JEV), Yellow Fever Virus(YFV) and Hepatitis C Virus (HCV). Sequences of Dengue virus, West Niles Virus and Hepatitis C Virus were obtained from RCSB Protein Data Bank (PDB) database. Meanwhile, sequences of Japanese Encephalitis Virus and Yellow Fever Virus were obtained from National Centre for Biotechnology Information (NCBI) database.

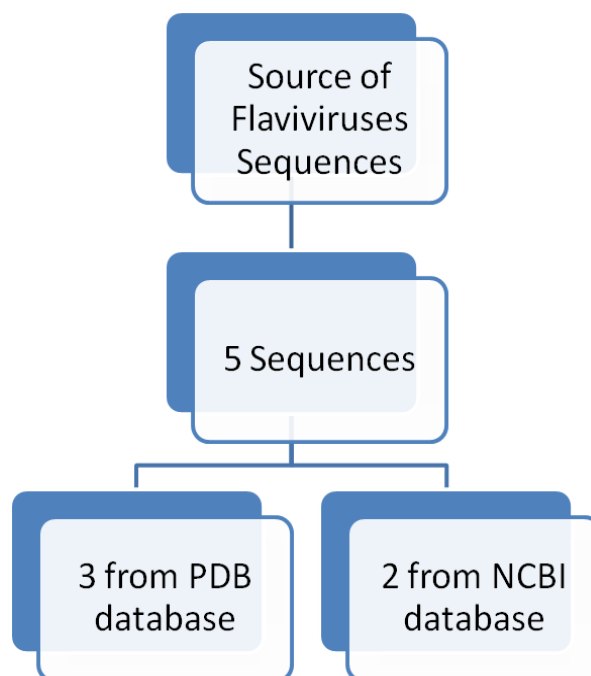


Figure 3.1 Source of Flaviviruses Sequences

3.2 Multiple Sequence Alignment of Flaviviruses

The sequences of the involved flaviviruses were downloaded in FASTA format from PDB and NCBI database. The sequences used in this study are shown in Table 3.1.

Table 3.1: Five species of flaviviruses sequences retrieved from PDB or NCBI database.

SPECIES	Data base	PDB ID/ Accession No.	FASTA SEQUENCE (PROTEIN)
YELLOW FEVER VIRUS (YFV)	NCBI	NP_776005	>DGLELKKLGEVSWEEEEAEISGSSARYDVALSEQ GEFKLLSEEKVPWDGGGGSGGGSGDVLWDIPTP KIIIECEHLEDGIYGFQSTFLGASQRGVGVAQG GVFHTMWHVTRGAFVLRNGKKLIPSWASVKEDLV AYGGSWKLEGRWDGEEVQLIAAVPGKNVNVVQT KPSLFKVRNGGEIGAVALDYPSGTSGSPIVNRNG EVIGLYNGILVGDNSFVSAISQTEVKEEGKEEL QEIPTMLKK
JAPANESE ENCEPHALITIS VIRUS (JEV)	NCBI	NP_775670	>TDMWLERAADISWEMDAAITGSSRRLDVKLDDD GDFHLIDDPGVPWKGGGGSGGGGGVFWDTSPSPK PCSKGDTTGTGYRIMARGILGTYQAGVGVMYENV FHTLWHHTTRGAAIMSGEGKLTYPWGSVKEDRIAY GGPWRFRDRKWNGTDDVQVIVVEPGKAAVNIQTKP GVFRTPFGEVGAVALDYPRGTSGSPILDSNGDII GLYNGVELGDGSYVSAIVQGERMEEPVPPEAYT
WEST NILES VIRUS (WNV)	PDB	2FP7	>GDTTGTGYRIMTRGLLSYQAGAGVMVEGVFHT LWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGP WKLQHKWNGHDEVQMIIVVEPGKNVKNVQTKPGVF KTPEGEIGAVTLDYPTGTSGSPIVDKNGDVIGLY NGVIMPNGSYISAIVQGERMEEPAPAGFEPEML RKK
HEPATITIS C VIRUS (HCV)	PDB	4I31	>MAPITAYSQQTRGLLGCIITSLTGRDKNQVDGE VQVLSTATQSFLATCVNGVCWTVYHGAGSKTLG PKGPITQMYTNVDQDLVGPAPPGARSMTPCTCG SSDLYLVTRHADVIPVRRRGDSRGSLLSPRPVSY LKSSGGPLLCPHGHVVGIFRAAVCTRGVAKAVD FIPVESMETTMRASKKKK
DENGUE VIRUS (DENV)	PDB	2FOM	>AGVLWDVPSPPVVGKAELEDGAYRIKQKILGY SQIGAGVYKEGTFHTMWHVTRGAVLMHKGKRIEP SWADVKKDLISYGGGWKLEGEWKEGEEVQVLAL PGKNPRAVQTKPGLFKTNTGTIGAVSLDFSPGTS GSPIVDKKGVVGLYNGVVTTRSGAYVSAIANTE KSIEDNPEIEDDIFRK

The alignment of these sequences was obtained by using Clustal Omega tool. Clustal Omega is a multiple sequence alignment program for proteins. Biologically meaningful multiple sequence alignments of divergent sequences are produced by using this tool. The steps involved in this method are shown in Figure 3.2.

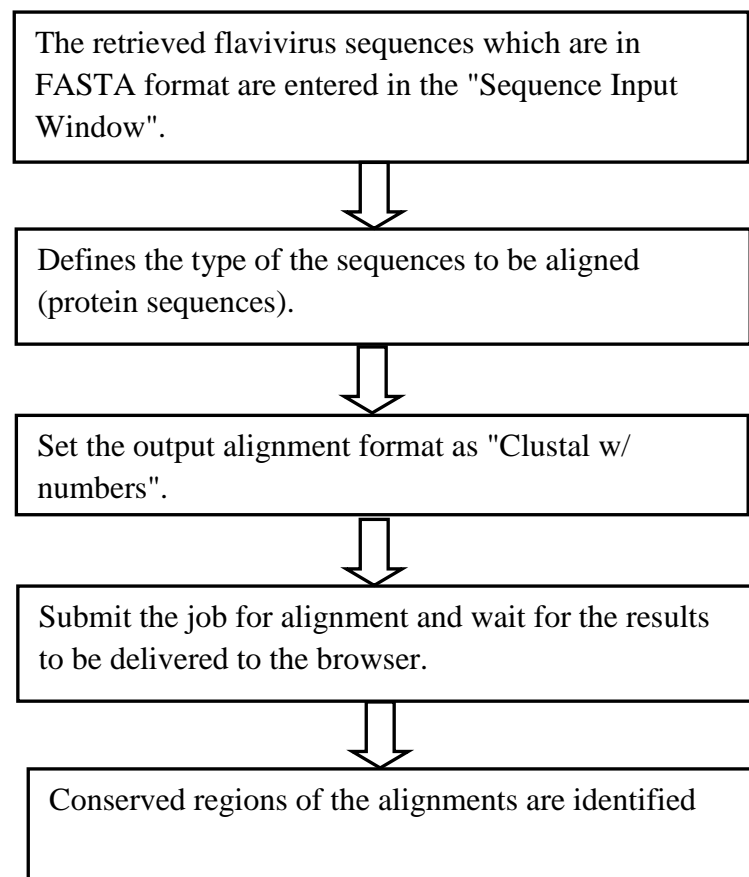


Figure 3.2: Flow chart aligning protein sequences by using Clustal Omega Online tool.

3.3 Identification of anti-viral target binding site

RCSB PDB (Protein Data Bank) database is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies. These information help students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease. The protein structures available in this database are essential in antiviral research field.

Identification of the binding sites in the involved protein structures are identified using literature research of the journals that report about the structures of the viruses. The mechanism of finding the binding sites through literature research is shown in Figure 3.3.

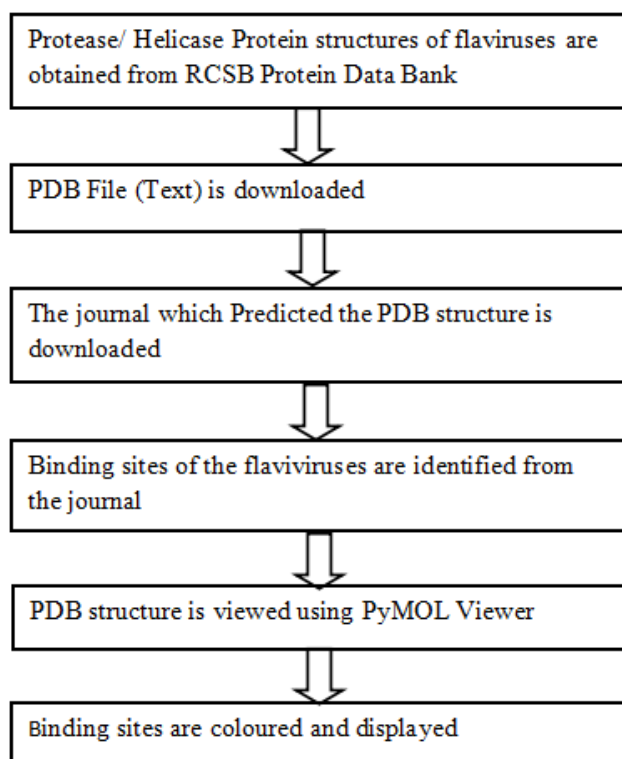


Figure 3.3: Flow chart of identifying protease binding sites through literature research.

3.3.1 Identification of Binding Sites Using PyMOL Viewer

PyMOL Viewer is a software package used for customization of 3D biomolecular images to provide users with precise and powerful control. Binding sites of the proteins at serine protease is coloured using tools in PyMOL based on the findings acquired from multiple sequences alignment and literature research. The binding sites of the protein structure are identified using PyMOL viewer and viewed in several forms to show the binding sites clearly (Figure 3.4).

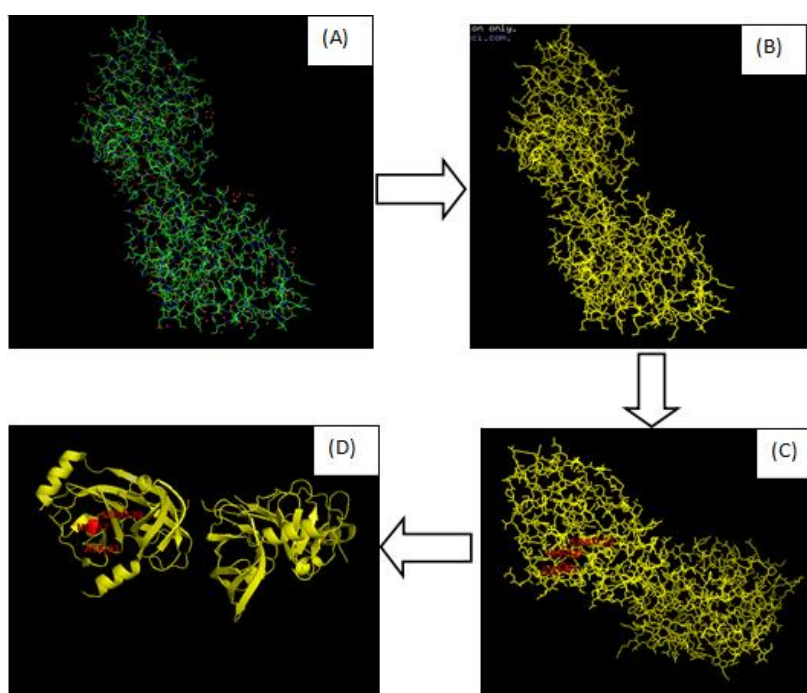


Figure 3.4: Steps involved in identifying binding sites using PyMOL Viewer. The structure shown in the figure is for Hepatitis C Virus (Protein Data Bank Accession code 4I31) (A) Crystal structure of the flaviviruses species. (B) Water molecules are removed from the structure and the colour of the structure is changed to one colour; yellow for easy observation. (C) The residues in the structure are labeled and the preferred residues which are the possible binding sites are chosen. The chosen residues

are coloured red and labeled. Label of other residues are hide. (D) The structure with labeled binding sites is viewed in different form (eg: Ribbon view)