WORLD JOURNAL OF PHARMACEUTICAL RESEARCH
SJIF Impact Factor 8.084
Volume 9, Issue 3, 555-570. Research Article ISSN 2277– 7105

# IN SILICO APPROACH TO STUDY SEQUENCE ALIGNMENT, PREDICTION OF PRIMER AND HOMOLOGY MODELLING

**Nikunj Patel[1], Nayan Prajapati[2], Richa Patel[3], Khushi Patel[4] and Haresh Kalasariya***

SMT. S. S. Patel Nootan Science & Commerce College, Sankalchand Patel University, Visnagar-384 315.

Article Received on
06 Jan. 2020,

Revised on 26 Jan. 2020,
Accepted on 16 Feb. 2020,

DOI: 10.20959/wjpr20203-16919

***Corresponding Author**
**Haresh Kalasariya**
SMT. S. S. Patel Nootan Science & Commerce College, Sankalchand Patel University, Visnagar-384 315.

## ABSTRACT

The aim of present study is to check the relationship between different nucleotide sequences of viruses belongs to Flaviviridae family retrieved from NCBI Database. The studied parameters include Sequence alignment, prediction of condensed primers and Homology modeling construction. The Clustal series of programs are widely used in molecular biology for the multiple sequence alignment and for constructing phylogenetic trees from nucleotide sequences. Choosing appropriate primers is probably the single most important factor affecting the polymerase chain reaction (PCR) process. Primer-BLAST offers adaptable choices to change the specificity threshold and other properties. This tool is useful to predict conserved primers in nucleotide query sequence. Three-dimensional structure of protein predicted by using Homology modeling study. This homology modeling prediction performed by a webtool Swiss model that uses amino acid sequences as an input query in workspace. This modeling study provides valuable information about 3D structure of models and their target template sequence alignment as well model coordination. In this pursuit, all the prediction studied parameters proceeds in silico.

**KEYWORDS:** Homology modeling, NCBI database, Primer-BLAST, Sequence alignment.

## INTRODUCTION

The genus of Flaviviridae family, Flavivirus possesses Single-stranded +ve sense RNA. This genus is arthropod-borne and often link between arthropod vectors () and vertebral animal hosts.[1] The genus Flavivirus comprises many important viruses such as Yellow fever virus, tick-borne encephalitis, zika virus, West Nile virus, apoi virus, palm creek virus, Dengue

virus, and Parramatta river virus.[2,3] Among them, yellow fever virus, Zika virus, Edge hill virus and apovirus are targeted for this present focus, these flaviviruses transmitted by the bite of the infection of female mosquitoes *Aedes aegypti*. Common symptoms associated with the infection of this virus include headache, fever, nausea vomiting, fatal, heart, liver and kidney disease, rash, joint pain, and red eyes, fatigue, loss of appetite etc.[4]

Sequence alignment is the best way to assess the relationship between query sequences. This process is helpful to count the percentage of nucleotide, identical and similar nucleotides in terms of % of sequence identity & sequence similarity.[5] Among them, many tools or software are useful to study sequence alignment such as Clustal Omega, Kalign, MAFFT, MUSCLE, MView, T-Coffee, etc. Among them, Clustal Omega is one of the useful additions to the Clustal family. It is a multiple sequence alignment program that uses techniques to generate alignments between three or more sequences. It helps to study biologically important multiple sequence alignments of divergent sequences. The phylogenetic relationships can be view in the form of Cladograms or Phylograms.[6,7]

Including different types of sequence alignment, MSA is the technique with the advantages of carrying alignment between multiple nucleotide sequences.[8] Primer is the sequence of RNA where DNA polymerase can bind and initiated its replication. Due to prime importance of primer in PCR reaction or amplification process, selection of best suitable and template specific primer is the first step in need. Here in this study, Primer-BLAST is helped to predict available primer in searched queries.[9]

PCR is a common method for DNA multiplication or DNA amplification applicable in many fields such as genetic diagnosis, biomedical testing, molecular research as and forensic study, etc. The PCR product affected by many other conditions such as components of PCR mixtures, DNA template, reaction conditions, concentration of enzymes as well as template specific primer. Among them, Primer designing is one of the critical factors.[10,11]

Homology modeling is the method to generate a reliable, accurate and precise three-dimensional structure of protein.[12,13] Protein characterization prediction, shared region of homology modeling[14], determination of consequence, prediction of secondary and tertiary structures of new query sequence, molecular evolution study, phylogenetic study[8] and construction of cluster hierarchy by this techniques.[15] The output of homology modeling display 3D structure of model and their target template sequence alignment. It also shows the

model coordination details and sequence features/scoring schemes for better understanding of 3D molecular view.[16,17]

The objective of the present study is to check the sequence similarities between targeted nucleotide sequences by multiple sequence alignment. Primer prediction study carried out by using primer BLAST. And further, selected nucleotide sequences converted into amino acid sequences by using EMBOSS Transeq. And lastly, homology modeling study carried out by using amino acid sequences as a query in SWISS Model.

## SEQUENCE RETRIEVAL FROM NCBI

A total of 4 different Flavivirus have selected from NCBI (National Center for Biotechnology Information) database for this study. The accession id of this Selected query are S71025.1 (Yellow fever virus envelope protein (E) gene, partial cds), KX893855.1 (Zika virus/Homo sapiens/VEN/UF-2/2016), AF275877 (Edge Hill virus strain P1553 NS5 protein gene, partial cds) and NC_003676 (Apoi VirusAfter query searched in nucleotide database, FASTA tool can be observed in the obtained output (https://www.ncbi.nlm.nih.gov/nuccore/?term=). The nucleotide sequences were obtained in the FASTA file format by entering accession id as an input in FASTA software.[18]

## SEQUENCE ALIGNMENT

A sequence alignment study carried out by using BLAST. BLAST (Basic Local Alignment Searching Tool) is an algorithm-based tool for finding out the regions of similarities or differences. The functional and evolutionary relationship between targeted sequences and to find member of gene families can be done by BLAST. The alignment study of DNA, RNA or protein sequence can be carried out by using different web BLAST such as Blastn, tblastn, Protein BLAST respectively. Among them, Blastn (nucleotide BLAST) that is used to compare four nucleotide sequence with each other. Blastn can be accessed on https://blast.ncbi.nlm.nih.gov/Blast.cgi. Here selected nucleotide sequence or downloaded FASTA sequence file used as an input to complete alignment study prediction.[19,20]

## PRIMER PREDICTION

Nowadays, there are many tools useful for this purpose such as Primer-BLAST, Quant Prime, NetPrimer, AutoPrime, SFOLD, RF-Cloning, etc. In which, the Primer-BLAST program is run on a farm of machines at the NCBI to provide faster and more reliable service to users. In order to expand the chance of discovering specific primer pairs, at least one primer (for a

given primer pair) ought to be situated in regions where the PCR template. The Primer-BLAST program is useful to check the target specificity of the generated primer pairs. It is a combination of Primer3 and BLAST. The former one is used to generating the candidate primer pairs for a given template sequence. And the latter one is used to check specificity of matches between the primers and targets. Primer-BLAST is a tool to find specific primers to PCR template. Same as BLAST selected nucleotide sequence or downloaded FASTA sequence used as an input in this tool. This tool can be accessed on https://www.ncbi.nlm.nih.gov/tools/primer-blast/.[21]

## HOMOLOGY MODELLING STUDY

Homology modelling requires amino acid or peptide sequence as a query input. EMBOSS (European Molecular Biology Open Software Suite) is a huge collection of tools for molecular biology. Among all the tools, EMBOSS Transeq is a helpful program to translate nucleotide sequences into their corresponding peptide sequences.[22]

This homology modeling prediction was performed on the webserver SWISS-MODEL (https://swissmodel.expasy.org/). Protein homology modeling can be prepared at different levels of complexity, tools for template selection, model construction, and build structure quality evaluation.[23] Different parameters such as Oligo state, Ligands, GMQE, QMEAN, C β, Solvation, Torsion, Template, Sequence identity, Protein description used to build homology modeling of searched sequence.[24,25]

## SEQUENCE ALIGNMENT STUDY

In this study, selected nucleotide sequences for different viruses downloaded from nucleotide database NCBI. This sequences with its accession id and no. of total base pairs tabulated in table no. 1. In alignment study, the output of Clustal W showed multiple sequence alignment between sequence no. 1 to 4. Predicted alignment score between each pair tabulated in table no. 2. The maximum alignment score was 63% found between sequences 2 and 4(2:4) and the minimum score was 4% found between sequences 2 and 3(2:3). Whereas the total alignment score was 33185 that predicted by ClustalW.

## PRIMER PREDICTION

Primer BLAST predicts basic characteristics such as Length of total base sequence, start and stop (initiation and termination) number of frame, melting temperature (Tm), GC content, self complementarity, etc. The outcome for searched query sequences tabulated in table no. 3-

6.

**HOMOLOGY MODELLING**

A comparative modelling study of proteins was carried out by using web server tool Swiss Model. In this study, it requires amino acid sequence or peptide sequence as an input. This query sequence obtained by translating nucleotide sequence of selected virus strains by using EMBOSS Transeq web tool. In model evaluation, output revealed different parameters such as GMQE, QMEAN, Oligo state, Ligands, and comparison plot. This model also expressed data about Global quality estimate and local quality estimate that recorded in table no. 7. Homology modeling prediction for all targeted sequences reported in model no. 1-4.

GMQE also is known as Global Model Quality Estimation, is quality estimation that Combines properties of the target-template alignment and the template search method. Out of 4 sequences, 3$^{rd}$ sequence showed higher scoring value compared to others. Composite estimator QMEAN worked on the basis of different geometrical properties and provides both global and local absolute quality estimation on the basis of one single model. QMEAN Z-score provides an estimate of the degree of nativeness in the model on a global scale. QMEAN Score for sequence 3$^{rd}$ found -1.56, which showed good agreement between model structure and experimental structures of similar size.

In the model output, the white area in the bar plots indicates a numerical value close to zero that indicates the property is similar with experimental structure of similar size. The individual Z-scores interaction potential between cβ atoms only, all atoms, the solvation potential, and the torsion angle potential. The found estimation represented in figure no. 1-4.

In local quality plot, each residue of the model reported on x-axis and the expected similarity to the native structure reported on the y-axis. In the output sequence number 1, 2 and 4 indicated the residues showed a score value below 0.6, that expected to be of low quality whereas, in case of sequence 3, most of the residues showed scoring value above 0.6 that suggest being of good quality. In the comparison plot, model quality scores of individual models are related to scores obtained for experimental structures of similar size. The x-axis shows protein length (number of residues). The y-axis is the normalized QMEAN score. Experimental protein structure represented by every dot. The obtained data was recorded in figure no. 5-8. In the output, red star represented the actual model. Sequence 2 & 4 found light grey that shows experimental structure is even further from the mean. Whereas sequence

1 & 3 found grey that indicates experimental structures with a |Z-score| between 1 and 2.

The finding of the present study would help in broadening understanding of different Flaviviruses such as Zika virus, Yellow fever virus, Edge Hill virus, and Apoi Virus, in particular, understanding the sequence homology study, evolutionary relationship study, as well as protein prediction by homology modeling for further research perspectives and purposes, which is as yet not an established phenomenon. This study reflects expected accuracy and reliability of the quality estimation as well as better quality estimation of entire structure and per residue. Recently, new features of the ClustalW and Swiss model are also being uploaded which can be used for further research prediction analysis.

**Table 1: Selected Nucleotide Sequences from NCBI Database.**

| Sequence No. | Nucleotide sequences | Accession ID | No. of Base Pairs |
|---|---|---|---|
| 01 | Apoi Virus | NC_003676 | 10166 |
| 02 | Edge Hill virus strain P1553 NS5 protein gene, partial cds | AF275877 | 986 |
| 03 | Yellow fever virus envelope protein (E) gene, partial cds | S71025.1 | 348 |
| 04 | Zika virus/Homo sapiens/VEN/UF-2/2016 | KX893855.1 | 10808 |

**Table 2: Multiple Sequence Alignment Study.**

<table>
<tr><td colspan="2">Selected nucleotide sequences for alignment<br>CLUSTAL 2.1 Multiple Sequence Alignments, Sequence format: Pearson</td></tr>
<tr><td colspan="2">Sequence 1: NC_003676.1 10116 bp<br>Sequence 2: AF275877.1 986 bp<br>Sequence 3: S71025.1 348 bp<br>Sequence 4: KX893855.1 10808 bp</td></tr>
<tr><td colspan="2">Alignment score for different pair of sequence</td></tr>
<tr><td>Sequences (1:2) Aligned. Score: 46<br>Sequences (1:3) Aligned. Score: 33<br>Sequences (1:4) Aligned. Score: 31<br>Sequences (2:3) Aligned. Score: 4<br>Sequences (2:4) Aligned. Score: 63<br>Sequences (3:4) Aligned. Score: 52</td><td>There are 3 groups<br>Group 1: Sequences: 2 Score:4125<br>Group 2: Sequences: 2 Score:12049<br>Group 3: Sequences: 4 Score:14123<br>Alignment Score 33185</td></tr>
</table>

**Table 3: Output of Primer-Blast for Sequence 4 (NC_003676).**

| No | Accession No | Number of primers | No | Length | Start | Stop | Tm | GC% | Self complementarity | Self 3'complementarity |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 Apoi Virus | NC_003676.1 | 10 | 1 | 20 | 7886 | 7905 | 59.97 | 55.00 | 6.00 | 2.00 |
| | | | 2 | 20 | 8172 | 8153 | 60.03 | 55.00 | 4.00 | 2.00 |
| | | | 1 | 20 | 6859 | 6878 | 59.97 | 55.00 | 6.00 | 2.00 |
| | | | 2 | 20 | 7128 | 7109 | 59.97 | 50.00 | 3.00 | 0.00 |
| | | | 1 | 20 | 7840 | 7859 | 59.97 | 55.00 | 4.00 | 0.00 |
| | | | 2 | 20 | 8386 | 8367 | 59.96 | 55.00 | 4.00 | 1.00 |
| | | | 1 | 20 | 5502 | 5521 | 59.97 | 50.00 | 5.00 | 1.00 |
| | | | 2 | 20 | 6177 | 6158 | 59.97 | 55.00 | 2.00 | 2.00 |
| | | | 1 | 20 | 2901 | 2920 | 59.96 | 55.00 | 3.00 | 2.00 |
| | | | 2 | 20 | 3464 | 3445 | 59.97 | 55.00 | 3.00 | 0.00 |
| | | | 1 | 20 | 8859 | 8878 | 60.03 | 55.00 | 4.00 | 2.00 |
| | | | 2 | 20 | 9013 | 8994 | 60.04 | 55.00 | 3.00 | 1.00 |
| | | | 1 | 20 | 478 | 497 | 59.96 | 50.00 | 4.00 | 0.00 |
| | | | 2 | 20 | 1127 | 1108 | 60.03 | 55.00 | 3.00 | 0.00 |
| | | | 1 | 20 | 6158 | 6177 | 59.97 | 55.00 | 2.00 | 1.00 |
| | | | 2 | 20 | 7013 | 6994 | 59.96 | 60.00 | 7.00 | 3.00 |
| | | | 1 | 20 | 7835 | 7854 | 59.97 | 55.00 | 4.00 | 2.00 |
| | | | 2 | 20 | 8808 | 8789 | 60.04 | 55.00 | 2.00 | 1.00 |
| | | | 1 | 20 | 6996 | 7015 | 59.96 | 60.00 | 4.00 | 1.00 |
| | | | 2 | 20 | 7854 | 7835 | 59.97 | 55.00 | 4.00 | 0.00 |

**Table 4: Output of Primer-Blast for Sequence 3 (AF275877).**

| No | Accession No | Number of primers | No | Length | Start | Stop | Tm | GC% | Self complementarity | Self 3'complementarity |
|---|---|---|---|---|---|---|---|---|---|---|
| 2 Edge | AF275877 | 10 | 1 | 20 | 187 | 206 | 59.97 | 50.00 | 4.00 | 0.00 |
| | | | 2 | 20 | 943 | 924 | 59.97 | 55.00 | 2.00 | 1.00 |
| | | | 1 | 20 | 3633 | 3652 | 59.97 | 55.00 | 3.00 | 2.00 |
| | | | 2 | 20 | 3717 | 3698 | 60.03 | 55.00 | 4.00 | 3.00 |
| | | | 1 | 20 | 3257 | 3276 | 60.04 | 55.00 | 6.00 | 2.00 |
| | | | 2 | 20 | 3652 | 3633 | 59.97 | 55.00 | 3.00 | 2.00 |
| | | | 1 | 20 | 649 | 668 | 59.97 | 55.00 | 5.00 | 2.00 |
| | | | 2 | 20 | 821 | 802 | 60.03 | 55.00 | 5.00 | 2.00 |
| | | | 1 | 20 | 7368 | 7387 | 60.03 | 55.00 | 4.00 | 1.00 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Hill Virus | | | 2 | 20 | 7651 | 7632 | 60.03 | 55.00 | 3.00 | 3.00 |
| | | | 1 | 20 | 10111 | 10130 | 59.96 | 55.00 | 6.00 | 2.00 |
| | | | 2 | 20 | 10954 | 10935 | 60.03 | 55.00 | 3.00 | 3.00 |
| | | | 1 | 20 | 5334 | 5353 | 59.96 | 60.00 | 4.00 | 1.00 |
| | | | 2 | 20 | 5917 | 5898 | 60.04 | 55.00 | 4.00 | 0.00 |
| | | | 1 | 20 | 26 | 45 | 60.04 | 50.00 | 7.00 | 1.00 |
| | | | 2 | 20 | 397 | 378 | 60.03 | 55.00 | 4.00 | 2.00 |
| | | | 1 | 20 | 1451 | 1470 | 60.04 | 55.00 | 6.00 | 0.00 |
| | | | 2 | 20 | 2370 | 2351 | 60.03 | 55.00 | 3.00 | 2.00 |
| | | | 1 | 20 | 7790 | 4509 | 59.96 | 50.00 | 3.00 | 2.00 |
| | | | 2 | 20 | 5353 | 5334 | 59.96 | 60.00 | 4.00 | 3.00 |

**Table 5: Output of Primer-Blast for Sequence 1 (S71025.1).**

| No | Accession No | Number of primers | No | Length | Start | Stop | Tm | GC% | Self complementarity | Self 3' complementarity |
|---|---|---|---|---|---|---|---|---|---|---|
| 3 Yellow fever | S71025 | 10 | 1 | 20 | 1701 | 1720 | 59.97 | 55.00 | 5.00 | 2.00 |
| | | | 2 | 20 | 2523 | 2504 | 59.97 | 55.00 | 5.00 | 3.00 |
| | | | 1 | 20 | 1929 | 1948 | 60.03 | 55.00 | 3.00 | 1.00 |
| | | | 2 | 20 | 2662 | 2643 | 60.03 | 55.00 | 6.00 | 3.00 |
| | | | 1 | 20 | 7353 | 7372 | 59.97 | 55.00 | 4.00 | 2.00 |
| | | | 2 | 20 | 8099 | 8080 | 60.03 | 55.00 | 4.00 | 2.00 |
| | | | 1 | 20 | 5771 | 5790 | 57.90 | 55.00 | 6.00 | 1.00 |
| | | | 2 | 20 | 6112 | 6093 | 60.03 | 60.00 | 5.00 | 2.00 |
| | | | 1 | 20 | 7692 | 7711 | 60.03 | 55.00 | 4.00 | 3.00 |
| | | | 2 | 20 | 8559 | 8540 | 59.97 | 55.00 | 6.00 | 2.00 |
| | | | 1 | 20 | 2753 | 2772 | 59.97 | 55.00 | 3.00 | 0.00 |
| | | | 2 | 20 | 3105 | 3086 | 60.03 | 55.00 | 6.00 | 1.00 |
| | | | 1 | 20 | 4200 | 4219 | 60.03 | 60.00 | 5.00 | 0.00 |
| | | | 2 | 20 | 4893 | 4874 | 60.03 | 55.00 | 5.00 | 3.00 |
| | | | 1 | 20 | 5775 | 5794 | 59.97 | 55.00 | 3.00 | 1.00 |
| | | | 2 | 20 | 6606 | 6587 | 59.96 | 60.00 | 6.00 | 2.00 |
| | | | 1 | 20 | 138 | 157 | 59.96 | 55.00 | 3.00 | 0.00 |
| | | | 2 | 20 | 295 | 276 | 60.03 | 55.00 | 2.00 | 0.00 |
| | | | 1 | 20 | 9899 | 9918 | 60.04 | 55.00 | 4.00 | 3.00 |
| | | | 2 | 20 | 10040 | 10021 | 60.04 | 55.00 | 3.00 | 3.00 |

**Table 6: Output of Primer-Blast for Sequence 2 (KX893855.1).**

| No | Accession No | Number of primers | No | Length | Start | Stop | Tm | GC% | Self complementarity | Self 3'complementarity |
|---|---|---|---|---|---|---|---|---|---|---|
| 4 Zika Virus | KX893855.1 | 10 | 1 | 20 | 409 | 428 | 59.97 | 55.00 | 2.00 | 0.00 |
| | | | 2 | 20 | 690 | 671 | 59.97 | 60.00 | 4.00 | 2.00 |
| | | | 1 | 20 | 7824 | 7843 | 59.97 | 55.00 | 4.00 | 3.00 |
| | | | 2 | 20 | 8653 | 8634 | 59.97 | 55.00 | 3.00 | 2.00 |
| | | | 1 | 20 | 3562 | 3581 | 59.97 | 50.00 | 3.00 | 0.00 |
| | | | 2 | 20 | 4273 | 4254 | 6004 | 55.00 | 4.00 | 1.00 |
| | | | 1 | 20 | 1463 | 1482 | 59.97 | 55.00 | 8.00 | 2.00 |
| | | | 2 | 20 | 2146 | 2127 | 60.04 | 50.00 | 3.00 | 2.00 |
| | | | 1 | 20 | 671 | 690 | 59.97 | 60.00 | 4.00 | 2.00 |
| | | | 2 | 20 | 1513 | 1494 | 59.97 | 55.00 | 5.00 | 0.00 |
| | | | 1 | 20 | 2127 | 2146 | 60.04 | 55.00 | 3.00 | 0.00 |
| | | | 2 | 20 | 2494 | 2475 | 59.97 | 50.00 | 4.00 | 0.00 |
| | | | 1 | 20 | 5431 | 5450 | 59.96 | 55.00 | 3.00 | 0.00 |
| | | | 2 | 20 | 5505 | 5486 | 60.04 | 55.00 | 4.00 | 0.00 |
| | | | 1 | 20 | 7326 | 7345 | 59.96 | 50.00 | 3.00 | 3.00 |
| | | | 2 | 20 | 7785 | 7766 | 59.97 | 50.00 | 4.00 | 0.00 |
| | | | 1 | 20 | 7326 | 7345 | 59.96 | 50.00 | 3.00 | 3.00 |
| | | | 2 | 20 | 7785 | 7766 | 59.97 | 50.00 | 4.00 | 0.00 |
| | | | 1 | 20 | 1034 | 1053 | 60.04 | 55.00 | 5.00 | 3.00 |
| | | | 2 | 20 | 1901 | 1882 | 60.04 | 60.00 | 4.00 | 0.00 |

**Table 7: Comparison of Homology Modeling Prediction For Four Query Sequences.**

| NO. | ACCESSION ID. | OLIGO STATE | LIGANDS | GMQE | QMEAN | Cβ | SOLAVTION | TORSION | TEMPLATE | SEQUENCE IDENTITITY | PROTEIN DESCRIPTION |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | NC_003676 | MONOMER | 2 X ZN | 0.19 | -1.84 | 1.42 | -0.17 | -1.50 | 6qsn.2.A | 55.29% | GENOME POLYPROTEIN |
| 2 | AF275877 | MONOMER | 1 X ZN | 0.81 | -3.30 | 0.70 | -2.53 | -2.35 | 6qsn.2.A | 69.91% | GENOME POLYPROTEIN |
| 3 | S71025.1 | MONOMER | NONE | 0.96 | -1.52 | 0.90 | -2.38 | -0.34 | 2jqm.1.A | 99.07% | ENVELOPE PROTEIN E |
| 4 | KX893855.1 | MONOMER | NONE | 0.02 | -4.25 | 1.85 | -2.00 | -3.24 | 6fv0.1.A | 17.19 | Kinesin light chain 1, Torsion-1A |

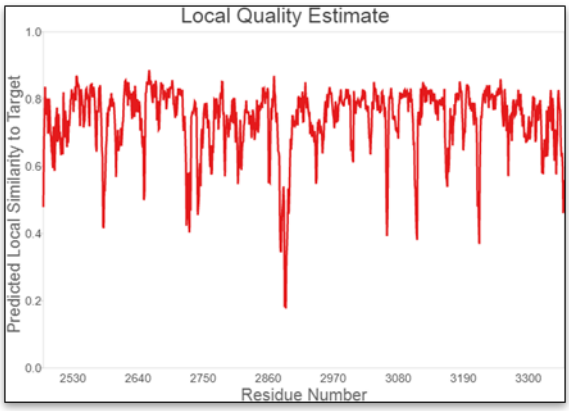



**Fig. 1: Local quality estimation for sequence 1.**

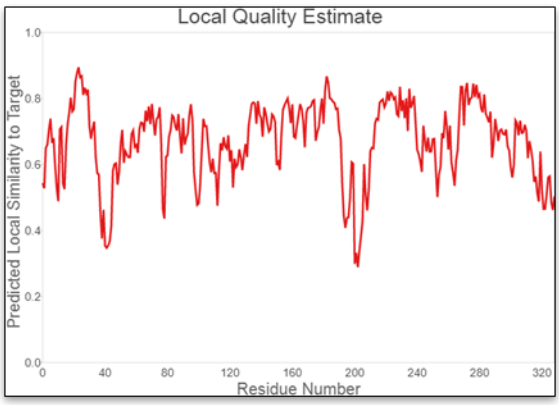



**Fig. 2: Local quality estimation for sequence 2.**

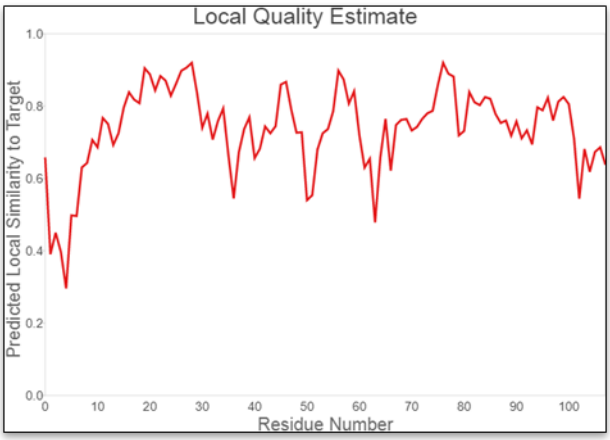



**Fig. 3: Local quality estimation for sequence 3.**

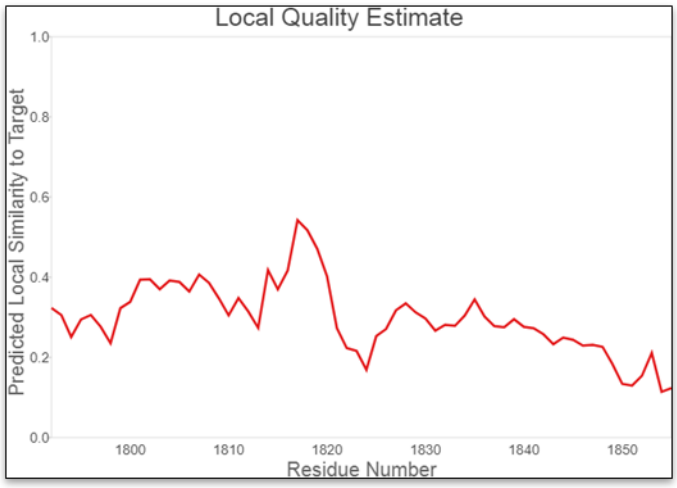



**Fig. 4: Local quality estimation for sequence 4.**

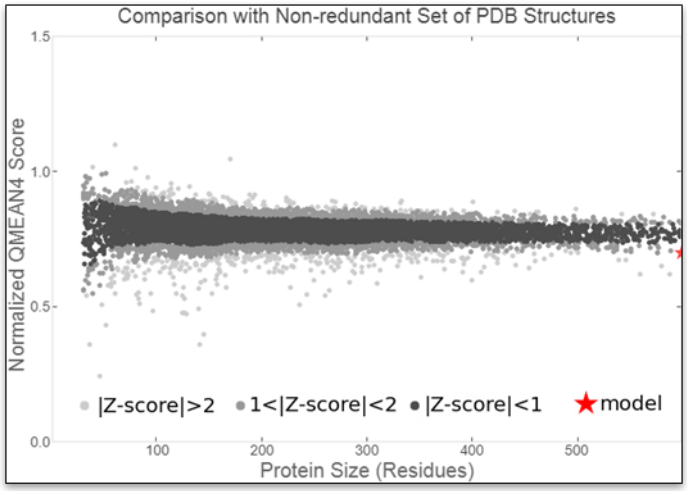



**Fig. 5: Comparison chart for sequence 1.**

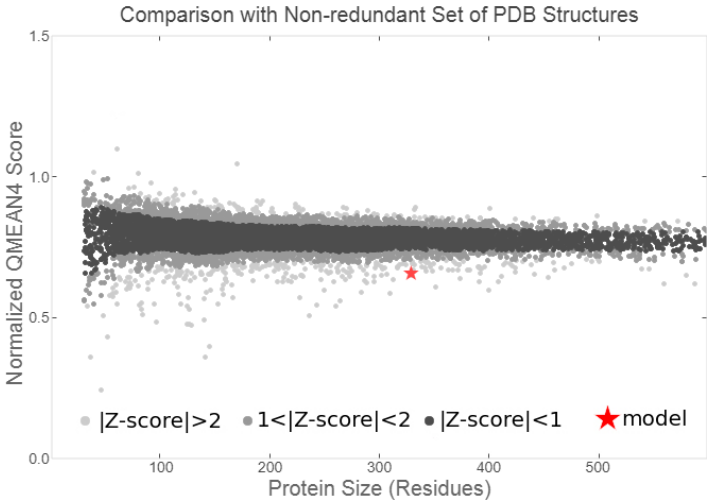



**Fig. 6: Comparison chart for sequence 2.**

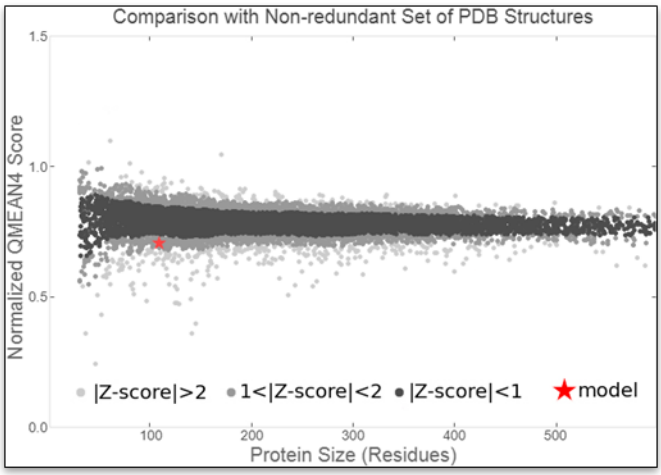



**Fig. 7: Comparison chart for sequence 3.**

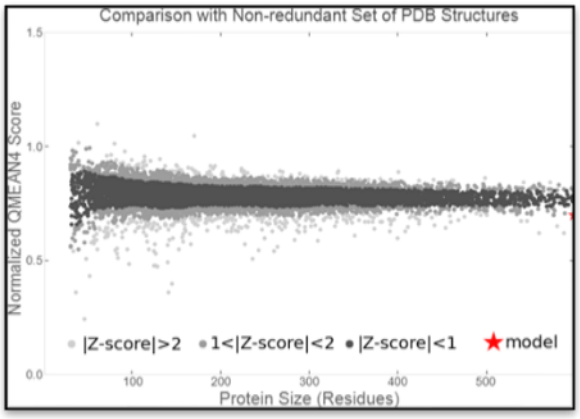
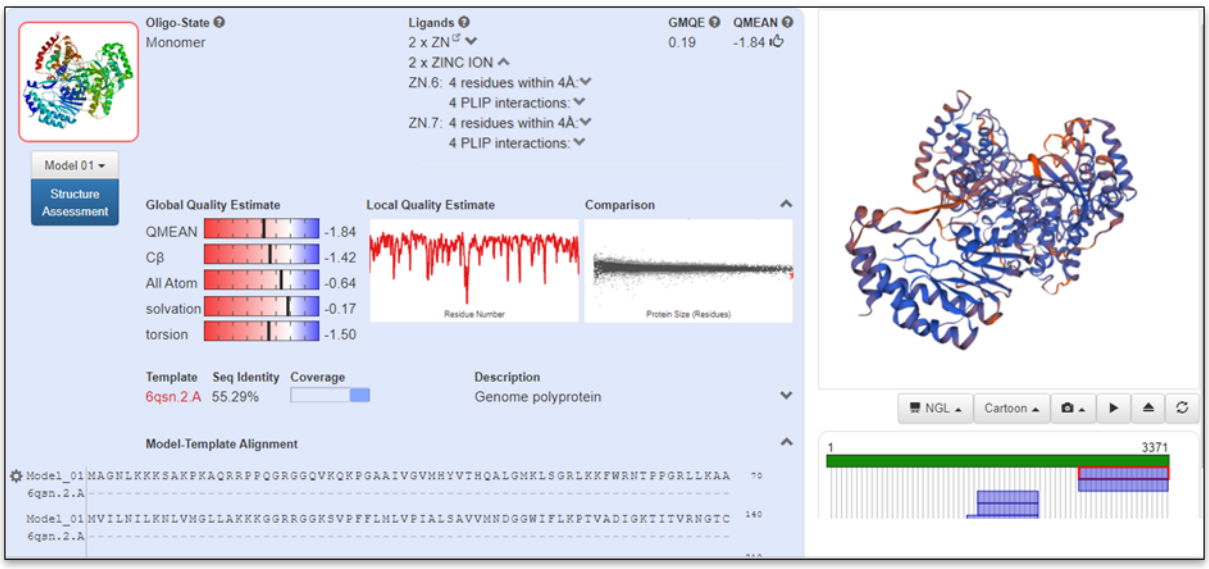
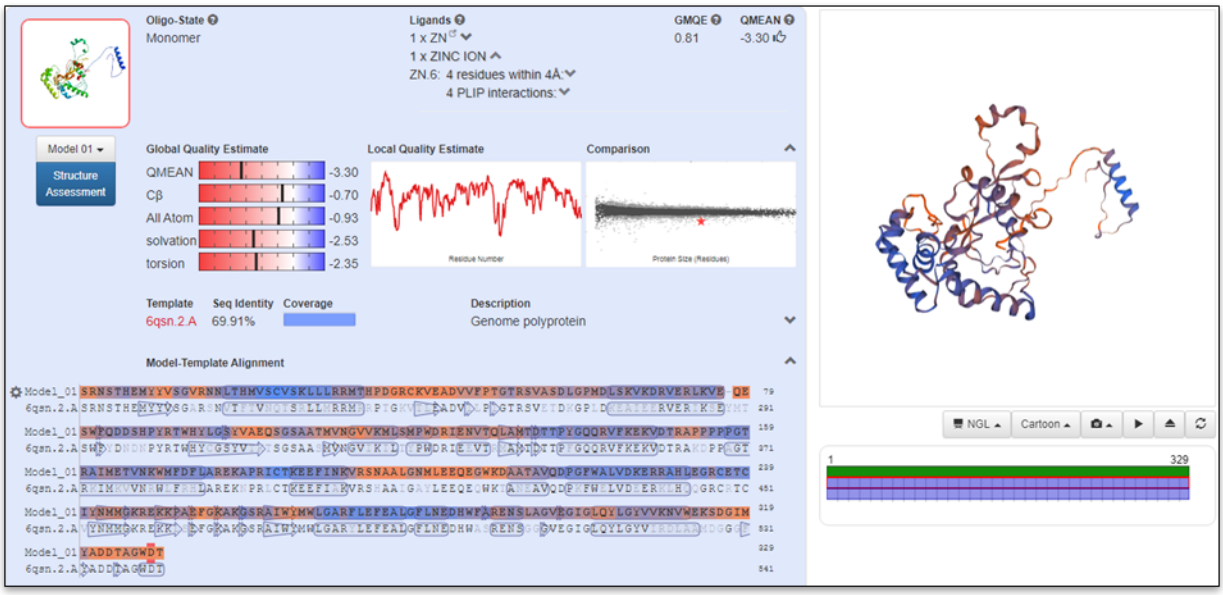


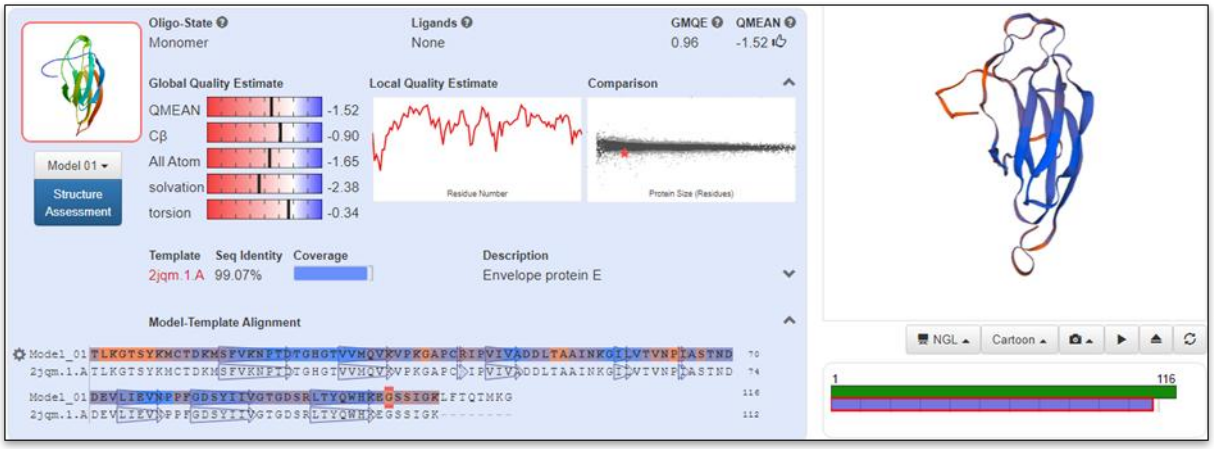

**Fig. 8: Comparison chart for sequence 4.**



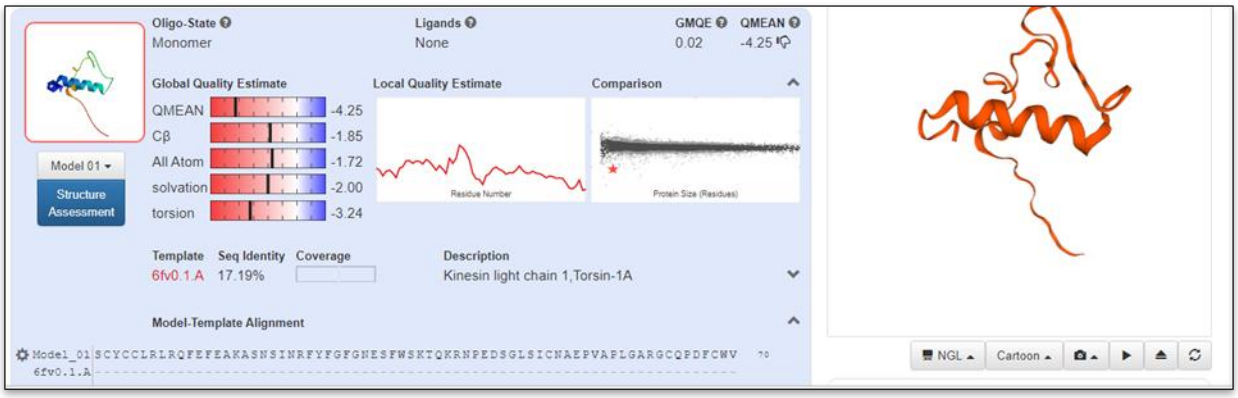

**Model 1: Homology modelling prediction report for NC_003676.**




**Model 2: Homology modelling prediction report For AF275877.**




**Model 3: Homology modelling prediction report for S71025.1.**




**Model 4: Homology modelling prediction report for KX893855.1.**

**ACKNOWLEDGMENTS**

I am very much thankful to the other co-authors for contributing and sharing their knowledge. As well as i am grateful to the Department of Microbiology and Sankalchand Patel University, Visnagar for providing facilities to complete this work.

**REFERENCES**

1. Leyssen P, De Clercq E, Neyts J. Perspectives for the Treatment of Infections with Flaviviridae. Clinical microbiology reviews, 2000 Jan 1; 13(1): 67-82.
2. Arruda LB, Mustafá YM, Meuren LM, Coelho SV. Pathways Exploited By Flaviviruses To Counteract The Blood Brain Barrier And Invade The Central Nervous System. Frontiers in microbiology, 2019; 10: 525.
3. Chepkorir E, Tchouassi DP, Konongoi SL, Lutomiah J, Tigoi C, Irura Z, Eyase F, Venter M, Sang R. Serological evidence of Flavivirus circulation in human populations in Northern Kenya: an assessment of disease risk 2016–2017. Virology journal, 2019 Dec; 16(1): 65.
4. Fernandez-Garcia MD, Mazzon M, Jacobs M, Amara A. Pathogenesis of flavivirus infections: using and abusing the host cell. Cell host & microbe, 2009 Apr 23; 5(4): 318-28.
5. McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic acids research, 2004 Jul 1; 32(suppl_2): W20-5.
6. Shugay M, Britanova OV, Merzlyak EM, Turchaninova MA, Mamedov IZ, Tuganbaev TR, Bolotin DA, Staroverov DB, Putintseva EV, Plevova K, Linnemann C. Towards error-free profiling of immune repertoires. Nature methods, 2014 Jun; 11(6): 653.
7. Syngai GG, Barman P, Bharali R, Dey S. BLAST: An introductory tool for students to Bioinformatics Applications. Keanean Journal of Science, 2013: 67-76.
8. Patil SS, Kumar V, Pai VR, Patil AK. Constructing phylogenetic tree and analysis using information retrieval approach for MYB tfr's of rice genome. In 2015 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE), 2015 Dec 19; (pp. 523-529). IEEE.
9. Potter KM, Dvorak WS, Crane BS, Hipkins VD, Jetton RM, Whittier WA, Rhea R. Allozyme variation and recent evolutionary history of eastern hemlock (Tsuga canadensis) in the southeastern United States. New Forests, 2008 Mar 1; 35(2): 131-45.
10. Abd-Elsalam KA. Bioinformatic tools and guideline for PCR primer design. african Journal of biotechnology, 2003; 2(5): 91-5.

11. Xiang Z. Advances in homology protein structure modeling. Current Protein and Peptide Science, 2006 Jun 1; 7(3): 217-27.
12. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. In The proteomics protocols handbook, 2005; (pp. 571-607). Humana press.
13. Wiederschain, G. Y. (2006). The proteomics protocols handbook. *Biochemistry (Moscow)*, *71*(6): 696–696. https://doi.org/10.1134/s0006297906060150
14. Münchau, W. (2017). Europe's four freedoms are its very essence. *The Future of Europe Project*, 808–824. https://doi.org/10.1110/ps.051892906.holdings.html
15. Vardhini, SR & Ekke S (2014). Homology Modelling and Structural Analysis of Her 2. *International Research Journal of Pharmacy*, *4*(12): 36–40. https://doi.org/10.7897/2230-8407.041208
16. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, and Higgins DG. (2007) Clustal W and Clustal X version 2.0. Bioinformatics, 23(21): 2947-8 PubMed: 17846036 DOI: 10.1093/bioinformatics/btm404
17. Shen J, Zhang W, Fang H, Perkins R, Tong W & Hong H. (2013). Homology modeling, molecular docking, and molecular dynamics simulations elucidated α-fetoprotein binding modes. *BMC Bioinformatics*, *14*(SUPPL.14). https://doi.org/10.1186/1471-2105-14-S14-S6
18. Pareek CS, Smoczynski R, & Tretyn A. (2011). Sequencing technologies and genome sequencing. *Journal of Applied Genetics*, *52*(4): 413–435. https://doi.org/10.1007/s13353-011-0057-x
19. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD. and Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol., 7: 539PMID: 21988835 DOI: 10.1038/msb.2011.75
20. Altschul, SF, Gish W, Miller W, Myers EW and Lipman DJ. (1990) Basic local alignment search tool. Journal of Molecular Biology, 215(3): 403-10 PubMed: 2231712
21. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, & Madden TL (2012). Primer-BLAST: a tool to design target-specific primers for a polymerase chain reaction. *BMC Bioinformatics*, *13:* 134. https://doi.org/10.1186/1471-2105-13-134
22. Rice P, Longden I and Bleasby A (2000) EMBOSS: The European Molecular Biology Open Software Suite Trends Genet, 16(6): 276-277 PubMed: 10827456

DOI: 10.1016/S0168-9525(00)02024-2.

23. Schwede T, Kopp J, Guex N, & Peitsch M C (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Research*, *31*(13): 3381–3385. https://doi.org/10.1093/nar/gkg520
24. Satyanarayana, S. D. V, Krishna, M. S. R., Pavan, P., & Jeereddy, S. (2018). Journal of Genetic Engineering and Biotechnology In silico structural homology modeling of nif A protein of rhizobial strains in selective legume plants. *Journal of Genetic Engineering and Biotechnology*, *16*(2): 731–737. https://doi.org/10.1016/j.jgeb.2018.06.006
25. Gasteiger, E, Gattiker A, Hoogland C, Ivanyi I, Appel R D, & Bairoch A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, *31*(13): 3784–3788. https://doi.org/10.1093/nar/gkg563