*Available online at* www.ijpab.com

DOI: http://dx.doi.org/10.18782/2320-7051.2103

**ISSN: 2320 – 7051**

*Int. J. Pure App. Biosci.* **3 (6):** 257-274 (2015)

International Journal of Pure & Applied Bioscience

*Research Article*

OPEN ACCESS

# *Insilico* Analysis of Mosaic Virus Proteins in Different Plants of Cucurbit Family

**Prashant Ankur Jain[1]*, Ankush Vinod Lal[1], Preetam Verma[2] and Ved Kumar Mishra[3]**

[1]Dept. of Computational Biology & Bioinformatics, Jacob School of Biotechnology & Bioengineering, Sam Higginbottom Institute of Agriculture, Technology & Sciences (Deemed to be University) Allahabad (U.P.)

[2]Dept. of MCE, Jacob School of Biotechnology & Bioengineering, Sam Higginbottom Institute of Agriculture, Technology & Sciences (Deemed to be University) Allahabad (U.P.)

[3]Ph.D. Scholar, Dept. of Computational Biology & Bioinformatics, Jacob School of Biotechnology & Bioengineering, Sam Higginbottom Institute of Agriculture, Technology & Sciences (Deemed to be University) Allahabad (U.P.)

*Corresponding Author E-mail: prash1721@gmail.com

## ABSTRACT

*Mosaic virus is caused by a variety of viruses which attack all members of the cucurbit family, but especially thrive on summer squash, cucumber and muskmelon plants. It is spread by a variety of methods and so is a serious disease for plants of the curcurbit family, including cucumbers, gourds, muskmelons, winter squash, summer squash, watermelons and pumpkins. The virus can infect cucumber, squash, muskmelon, and numerous other hosts in 34 plant families, including tomato, lima bean, beet, sweet corn, and sweet potato. Most often, actively growing and mature plants are affected. It rarely infects plants in the seedling stage, but will kill them quickly when it does. It causes a decrease in the number and the quality of the fruit. In this study with the help of various softwares we have elucidated the primary structure and secondary structure information and its components. The predicted and analyzed information can be further used in determining the three dimensional analysis and related studies for these proteins.*

***Key words:****Mosaic Virus, Poly Protein, p3 Protein, sequence analysis.*

## INTRODUCTION

Cucumber mosaic virus (CMV) causing viral diseases of many important plants worldwide have been isolated from pumpkin (*Cucurbitapepo* L.) plant leaves. Diseased plants had light green mottled foliage. Leaves were smaller than normal, yellow mottled and crinkled. Cucumber mosaic, caused by the cucumber mosaic virus, is one of the most widespread and destructive diseases on cucumber and muskmelon. The disease has been known since the early 1900's, and is now found worldwide. The virus can infect cucumber, squash, muskmelon, and numerous other hosts in 34 plant families, including tomato, lima bean, beet, sweet corn, and sweet potato. Most often, actively growing and mature plants are affected. It rarely infects plants in the seedling stage, but will kill them quickly when it does. It causes a decrease in the number and the quality of the fruit.

**Cite this article:** Jain, P.A., Lal, A.V., Verma, P. and Mishra, V.K., *Insilico* Analysis of Mosaic Virus Proteins in Different Plants of Cucurbit Family, *Int. J. Pure App. Biosci.* **3 (6):** 257-274 (2015). doi: http://dx.doi.org/10.18782/2320-7051.2103

Cucumber mosaic virus (CMV) has one of the broadest host ranges. CMV as a type species of the genus Cucumovirus in the family Bromoviridae is reported to infect 1287 plant species in 518 genera belonging to 100 families[3]. It is geographically widespread and has been reported in Europe, Asia, Australia, North America and India. In Lithuania, this virus is spread on black currant[13], leguminous[14], ornamental[12] and vegetable[16]plants, however not detected on pumpkin. The most common symptom induced by CMV is mosaic; however, severity of disease may range from no obvious symptoms in some crops to death of the host species. The virus causes fern leaf, stunting of vegetable crops and malformation of their fruits. It is transmitted by numerous species of aphid, through the sap, the seeds and dodder[2,5,7]. Morphologically CMV has rather characteristic about 30 nm polyhedral particles with hollow centre[9]. The genome consists of three plus sense single-stranded RNAs, packaged in separate particles. CMV particles contain about 18% RNA. The RNAconsists of 4 RNAs. Only largest RNA3 are required for infectivity[11]. The virions are not stable to freezing. Long-term storage of CMV is most reliable in the form of viral RNA, which is highly infectious, and very stable at −20ºC[10]. Great number of different CMV strains, serogroups, subgroups and biological variations has been described[1,4,6,8,15]. Unfortunately, there is no chemical control for mosaic virus, and plants need to be removed and destroyed promptly if they are infected with this viral disease. To control the spread of the disease by cucumber beetles and aphids, we need to control these insect populations with a diazinon containing insecticide repeating the application as much as necessary in seven day intervals. Although there are mosaic virus resistant cucumber varieties, so far no resistant varieties of muskmelon and summer squash are available to plant.Secondary structure in biochemistry and structural biology describes the general three-dimensional form of local segments of biopolymers such as proteins and nucleic acids (DNA/RNA). It does not, however, describe specific atomic positions in three-dimensional space, which are considered to be tertiary structure.

## MATERIALS AND METHODS

The target selected was tobacco mosaic virus protein. The primary sequence from retrieved from NCBI database. In the present work an integrated approach was used for sequence analysis. The query protein sequence of interest is subjected to an exhaustive sequence similarity search conducted over all accessible sequence databases by standard sequence analysis tools and it yields in 1D topology of the sequence. In the next step of this procedure "2D topology" is obtained which provides a platform to rationalize highly conserved sequence positions in terms of structural and functional relevance and further allows gaining qualitative insights into possible helix-helix interactions. The 2D topology is a prerequisite receptor sequence representation, defining the interrelation between the family-wide sequence characteristics.

Following Software and databases were used in this study.

**1.** Primary Sequence retrieval: NCBI (www.ncbi.nlm.nih.gov)
**2.** Sequence Alignment tools: BLASTP (www.ncbi.nlm.nih.gov/blast)
**3.** Primary Structure Analysis Protoparam: (http://an.expasy.org/tools/peotoparam.html)
**4.** Tm Region Prediction: TMHMM (www.cbs.dtn.dk/services/tmhmm-2.0)
**5.** Signal P3.0 Server Prediction: (www.cbs.dtn.dk/services/signalP)
**6.** Sumo Plot™ Prediction: (www.abgent.com/doc/sumoplot)
**7.** Secondary Structure Prediction: SOPMA:(www.cbs.dtn.dk/services/sopma)
**8.** Minimum System Configuration: Pentium 4 CPU, 3.8 GHz, 512 MB of RAM
**9.** Operating System: Windows

The sequence details were retrieved by searching GENE BANK using Entrees Browser. In predicting the 1D topology Proto Param was used for computing molecular weight, theoretical isoelectric point (p1), instability index and aliphatic index. Self Optimized prediction method (SOPM) was used to give a four state description of secondary structure (alpha helix, beta sheet, turn and coils).

The output width was set at 70. Prediction of protein membrances in the target was done using Tmpred & TMHMM. Prediction of cleavage sites as well as signal peptide/non-signal peptide prediction on combination of several rtificial neural networks and hidden markov models was done using SinalP server. Small Ubiquitin – like modifier (SUMO) protein can be conjugated to substrates by enzymes that operate in ubiquitylation, which mark proteins for rapid intercellular degradation. SUMOplot™ was used to

predict the probability for the SUMO consensus sequences (SUMO-CS) to be engaged in SUMO attachment of query sequence. Neutral network prediction for the attachment of O-linked N-acetylglucose amine (O-Glc NAC) in protein was done using Ying O Yang Server. By following all the above procedures, the protein sequence is subjected to an exhaustive sequence similarity search conducted over all accessible sequence databases by standard sequence analysis tools and yields its primary and secondary structure topology.

## RESULTS& DISCUSSIONS

- **Sequence Retrieval:**

Every step in sequence / structural analysis of the target proteins is found to be crucial to end - up with an accurate model . The accuracy and the reliability of a theoretical protein model often depend on the template structures and the alignment between the target and template sequence. The target sequences were taken from NCBI. To begin with, the FASTA formats of the proteins were retrieved.

>gi|51949951|ref|YP_077272.1| P3 protein [Watermelon mosaic virus]

GEAQQRMKCETALIKSIFKPKRMIQILEDDPYILLMGLISPSILIHMYRMKHFEKGIELWISK
EHSVAKIFIIMEQLTRKIAANDLLLEQLDIIAGTSQKLMDVLEDCPQSAHSYRTAKDLLAIY
IERRASNNQLIENGFVDINDQLYVTHEKIYVDRLKQEWHALSWLEKSSITWQLKRFTPHTE
QCLTKKVVEESSAYSRNFVSACFMNAQSHLKNVRNTFFRKCDQAWTASVRVLVRFIIATL
HKCYSDIVYLVNICLIFSLLVQMVSVLQGIVSTAKRDKAFVHMHKRIEDEQAVVHLYEMC
EKMENKHPSVEEFLSHVKKVRPELLPVAKSMTGQSEDVSAQ

>gi|365777336|gb|AEW91906.1| polyprotein, partial [Papaya ringspot virus W]

AAMIESWGYGELTHQIRRFYQWVLEQAPFNELARQGRAPYVSEVGLRRLYTSERGSMDE
LEAYIDKYFERERGDSPELLVYHESRIADDYQLVCSNNTHVFHQSKNEAVDAGLNEKLKE
KEKQKEKEKEKQKEKEKDDASDGNDVSTSTKTGERDRDVNVGTSGTFTVPRIKSFTDKMI
LPRIKGKTVLNLNHLLQYNPQQIDISNTRATQSQFEKWYEGVRNDYGLNDNEMQVMLNG
LMVWCIENGTSPDISGVWVMMDGETQVDYPIKPLIEHATPSFRQIMAHFSNAAEAYIAKR
NATERYMPRYGIKRNLTDISLARYAFDFYEVNSKTPDRAREAHMQMKAAALRNTSRRMF
GMDGSVSNKEENTERHTVEDVNRDMHSLLGMRN

>gi|343790821|dbj|BAK61797.1| 1a protein [Cucumber mosaic virus]

MATSSFNINELVASHGDKGLLATALVDKTAHEQLEEQLQHQRRGRKVYIRNVLGVKDSE
VIRNRYGGKYDLHLTQQEFAPHGLAGALRLCETLDCLDSFPSSGLRQDLVLDFGGSWVTH
YLRGHNVHCCSPCLGIRDKMRHAERLMNMRKIILNDPQQFDGRQPDFCTQPAADCKVQA
HFAISIHGGYDMGFRGLCEAMNAHGTTILKGTMMFDGAMMFDDQGVIPELNCQWRKIRS
AFSETEDVTPLSGKLNSTVFSRVRKFKTMVAFDFINESTMSYVHDWENIKSFLTDQTYSYR
GMTYGIERCVIHAGIMTYKIIGVPGMCPPELIRHCIWFPSIKDYVGLKIPASQDLVEWKTVR
ILMSTLRETEEIAMRCYNDKKAWMEQFKVILGVLSAKSSTIVINGMSMQSGERIDINDYHY
IGFAILLHTKMKYEQLGKMYDMWNASSISKWFAALTRPLRVFLSGVVHALFPTLRPREEK
EFLIKLSTFVTFNEECSFDGGEEWDVISSAAYVATQAVTDGKILAAQKAEKLAEKLAQPVI
EVSDSPEAPSQTPDDTAEVCGKEREVSELDSLSAQTRSPITRVAERATAMLEYAAYEKQLH
DTTVSNLKRIWNMAGGDDKRNSLEGNLKFVFDTYFTVDPMVNIHFSTGRWMRPVPEGV
VYSVGYNERGLGPKSDGELYIVNSECVICNSESLSTVTRSLQAPTGTISQVDGVAGCGKTT
AIKSIFEPSTDMIVTANKKSAQDVRMALFKSSDSKEACTFVRTADSVLLNECPTVSRVLVD
EVVLLHFGQLCAVMSKLKAVRAICFGDSEQIAFSSRDASFDMRFSKIIPDETSDADTTFRSP
QDVVPLVRLMATKALPKGTRSKYTKWVSQSKVKRSVTSRAIVSVTLVDLDPSRFYITMTQ
ADKASLISRAKEMNLPKTFWNERIKTVHESQGISEDHVTLVRLKSTKCDLFKQFSYCLVAL
TRHKVTFRYEYCGVLNGDLIAECVARA

>gi|37927249|gb|AAO45984.1| P3 protein [Zucchini yellow mosaic virus]

PYILLLGMISPTILVHMYRMRHFERGIEVWIKRDHEIGKIFVILEQLTRKVALAEVLVDQLN
LISEASPHLLEIMKGCQDNQRAYVPALDLLTIQVEREFSNKELKTNGYPDLQQTLFDMWE
KMYAKQLHNSWQELSLLEKSCVTVRLKQFSIFTERNLIQRAEEGKRASSLQ

Next, a set of related sequences were submitted to the multiple sequence alignment server CLUSTALW. The templates assigned were obtained through the BLASTp search using SWISS-PROT and PDB databases. The CLUSTAL's output format is compatible with GDE, PHYLIP or GCG packages. CLUSTALW involves a progressive strategy for aligning pairs of sequences .The CLUSTAL server was selected for sequence analysis as it exploits the fact that similar sequences are likely to be evolutionarily related and it expressed the degree of similarity in a relatively concised format . As part of its operation , the program produced information required to produce a phylogenetic tree.

**Fig.1:Cladogram of protein sequences**

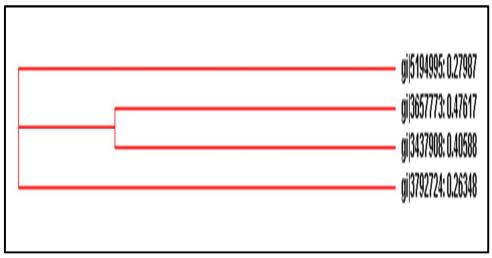
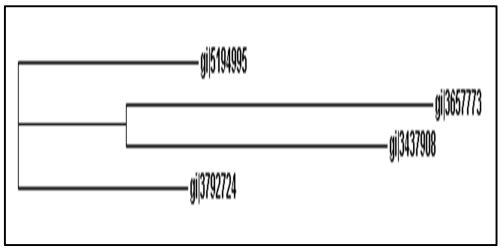



**Fig.2:Phylogram of protein sequences**




By the analysis of the phylogram and Cladogram it can be predicted that sequences gi/3657773 and gi/34379008 are forming a clade and more closely related to each other then other sequences.

- **Primary Structure Analysis (ProtParam): P3 protein (Watermelon mosaic virus)**

**Number of amino acids:**347 ;**Molecular weight:** 40288.0 ; **Theoretical pI:** 8.51

**Total number of negatively charged residues (Asp + Glu):** 41

**Total number of positively charged residues (Arg + Lys):** 45

**Atomic composition:**

| | | |
|---|---|---|
| Carbon | C | 1805 |
| Hydrogen | H | 2887 |
| Nitrogen | N | 489 |
| Oxygen | O | 511 |
| Sulfur | S | 21 |

**Formula:**$C_{1805}H_{2887}N_{489}O_{511}S_{21}$;**Total number of atoms:** 5713

**Instability index:**The instability index (II) is computed to be 45.08. This classifies the protein as unstable.

**Aliphatic index:**100.00 ;**Grand average of hydropathicity (GRAVY):** -0.144

- **Polyprotein, partial (Papaya ring spot virus W)**

**Number of amino acids:**390 ;**Molecular weight:** 45144.5 ; **Theoretical pI:** 6.01

**Total number of negatively charged residues (Asp + Glu):** 61

**Total number of positively charged residues (Arg + Lys):** 55

**Atomic composition:**

| | | |
|---|---|---|
| Carbon | C | 1961 |
| Hydrogen | H | 3074 |
| Nitrogen | N | 572 |
| Oxygen | O | 619 |
| Sulfur | S | 18 |

**Formula:**$C_{1961}H_{3074}N_{572}O_{619}S_{18}$ ;**Total number of atoms:** 6244

**Instability index:**The instability index (II) is computed to be 35.49. This classifies the protein as stable.

**Aliphatic index:**64.77 ;**Grand average of hydropathicity (GRAVY):** -0.864

- **1a protein (Cucumber mosaic virus)**

**Number of amino acids:**993 ;**Molecular weight:** 111421.8 ; **Theoretical pI:** 7.49

**Total number of negatively charged residues (Asp + Glu):** 118

**Total number of positively charged residues (Arg + Lys):** 119

**Atomic composition:**

| | | |
|---|---|---|
| Carbon | C | 4923 |
| Hydrogen | H | 7790 |
| Nitrogen | N | 1354 |
| Oxygen | O | 1472 |
| Sulfur | S | 60 |

**Formula:**$C_{4923}H_{7790}N_{1354}O_{1472}S_{60}$ ;**Total number of atoms:** 15599

**Instability index:**The instability index (II) is computed to be 42.32. This classifies the protein as unstable.

**Aliphatic index:**82.27 ;**Grand average of hydropathicity (GRAVY):** -0.206

- **P3 protein (Zucchini yellow mosaic virus)**

**Number of amino acids:**173 ;**Molecular weight:** 20327.7 ; **Theoretical pI:** 7.42

**Total number of negatively charged residues (Asp + Glu):** 22

**Total number of positively charged residues (Arg + Lys):** 22

**Atomic composition:**

| | | |
|---|---|---|
| Carbon | C | 915 |
| Hydrogen | H | 1468 |
| Nitrogen | N | 248 |
| Oxygen | O | 258 |
| Sulfur | S | 8 |

**Formula:**$C_{915}H_{1468}N_{248}O_{258}S_{8}$ ;**Total number of atoms:** 2897

**Instability index:**The instability index (II) is computed to be 46.05. This classifies the protein as unstable.

**Aliphatic index:**107.05 ;**Grand average of hydropathicity (GRAVY):** -0.218

- **Prediction of Theoretical pI/Mw:**

Compute pI/Mw is a tool which allows the computation of the theoretical pI (isoelectric point) and Mw (molecular weight) for a list of UniProt Knowledgebase (Swiss-Prot or TrEMBL) entries or for user entered sequences. The results are as following for our sequences:

**Table 1:Theoretical pI/Mw**

| Protein | Theoretical pI | Mw |
|---|---|---|
| P3 protein [Watermelon mosaic virus] | 8.51 | 40288.01 |
| Polyprotein, partial [Papaya ring spot virus W] | 6.01 | 45144.52 |
| 1a protein [Cucumber mosaic virus] | 7.49 | 111421.80 |
| P3 protein [Zucchini yellow mosaic virus] | 7.42 | 20327.71 |

- **InterPro Scan:**

InterPro Scan is a tool that scans given protein sequences against the protein signatures of the InterPro member databases, currently – PROSITE, PRINTS, Pfam, ProDom and SMART. The results are as following for our sequences

**Fig. 3: InterPro results of P3 protein [Watermelon mosaic virus]**

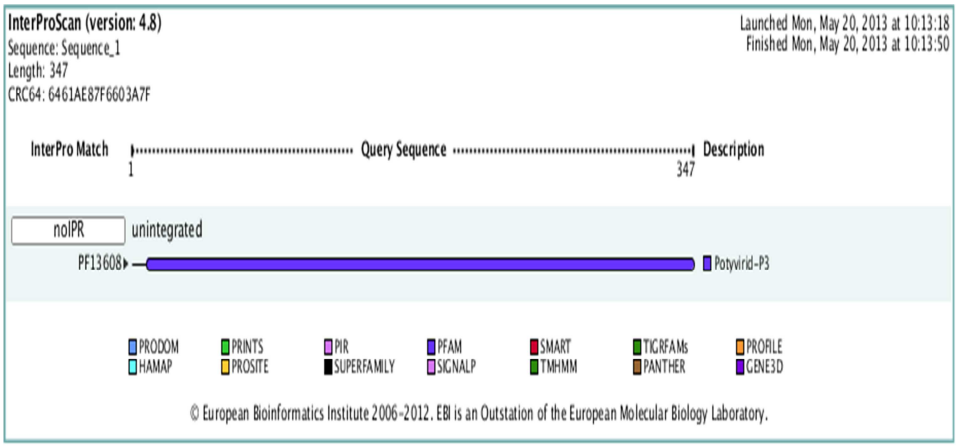



**Fig. 4: InterPro results of Polyprotein, partial [Papaya ring spot virus W]**

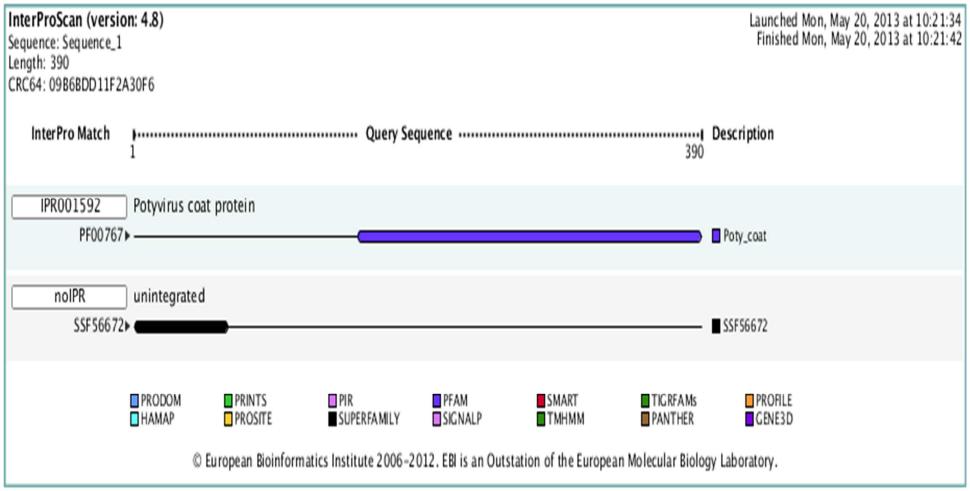


**Fig. 5: InterPro results of 1a protein [Cucumber mosaic virus]**

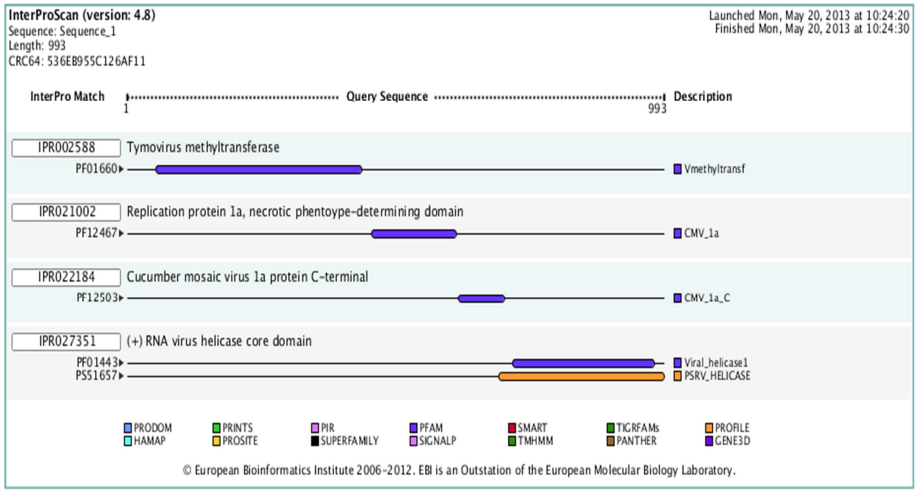
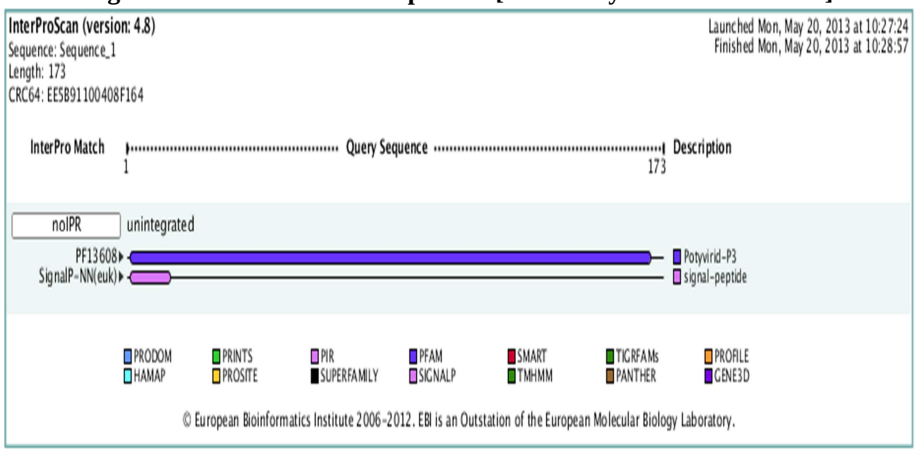



**Fig. 6: InterPro results of P3 protein [Zucchini yellow mosaic virus]**




SignalP was used to predict the presence of signal peptide sequence in the N –terminal region. There is no signal peptide predicted and the signal peptide probability is found to be zero. Thus, it could be safely predicted that the target sequences are non-secretory proteins .

**Fig.7: SignalP results of P3 protein [Watermelon mosaic virus]**

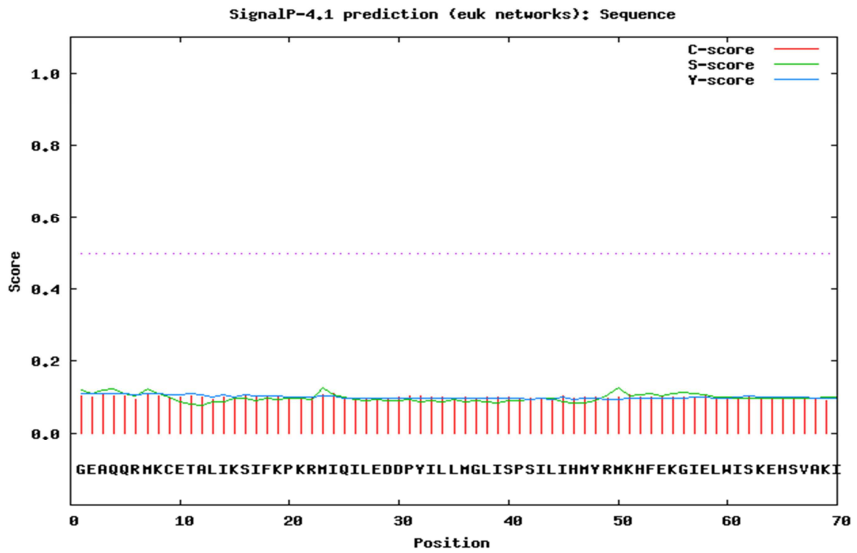


**Fig.8: SignalP results of Polyprotein, partial [Papaya ring spot virus W]**

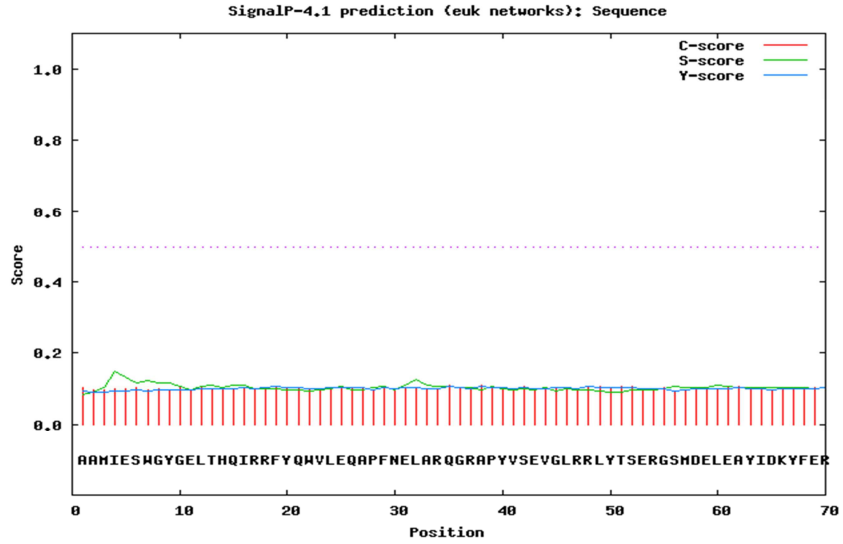



**Fig.9: SignalP results of 1a protein [Cucumber mosaic virus]**

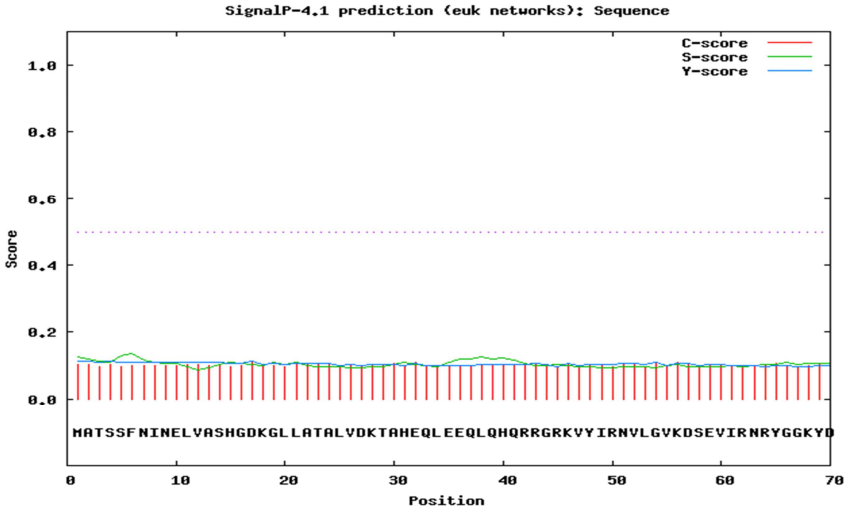
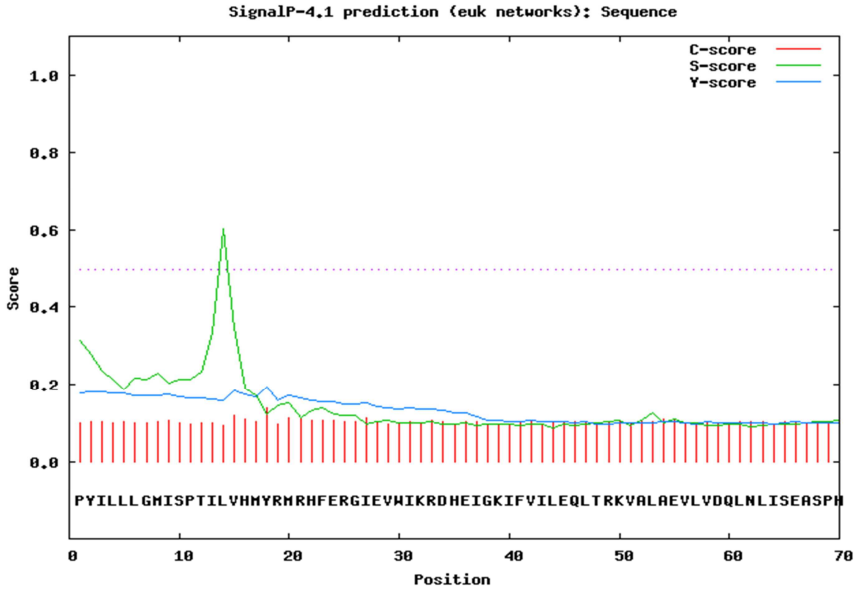



**Fig. 10: SignalP results of P3 protein [Zucchini yellow mosaic virus]**



- **The SUMOplot™ Analysis:**

Program predicts and scores sumoylation sites in protein. The SUMOplot™ Analysis Program predicts the probability for the SUMO consensus sequence (SUMO-CS) to be engaged in SUMO attachment. The SUMOplot™ score system is based on two criteria: direct amino acid match to SUMO-CS. substitution of the consensus amino acid residues with amino acid residues exhibiting similar hydrophobicity in the results the red colored amino acids represents Motifs with high probability, blue represents Motifs with low probability and green represents overlapping Motifs.

➢ **P3 protein [Watermelon mosaic virus]**

**Table 2. Motifs position in P3 protein [Watermelon mosaic virus]**

| No. | Pos. | Group | Score |
|---|---|---|---|
| 1 | K160 | IYVDR **LKQE** WHALS | 0.91 |
| 2 | K8 | EAQQR **MKCE** TALIK | 0.80 |
| 3 | K281 | GIVST **AKRD** KAFVH | 0.79 |
| 4 | K179 | SITWQ **LKRF** TPHTE | 0.56 |
| 5 | K308 | LYEMC **EKME** NKHPS | 0.50 |
| 6 | K51 | IHMYR **MKHF** EKGIE | 0.45 |
| 7 | K226 | RNTFF **RKCD** QAWTA | 0.44 |
| 8 | K69 | KEHSV **AKIF** IIMEQ | 0.44 |
| 9 | K284 | STAKR **DKAF** VHMHK | 0.15 |

➢ **Polyprotein, partial [Papaya ring spot virus W]**

**Table 3. Motifs position in Polyprotein, partial [Papaya ring spot virus W]**

| No. | Pos. | Group | Score |
|---|---|---|---|
| 1 | K172 | FTVPR **IKSF** TDKMI | 0.59 |
| 2 | K136 | KQKEK **EKDD** ASDGN | 0.50 |
| 3 | K66 | LEAYI **DKYF** ERERG | 0.15 |

➢ **1a protein [Cucumber mosaic virus]**

**Table 4. motifs position in 1a protein [Cucumber mosaic virus]**

| No. | Pos. | Group | Score |
|---|---|---|---|
| 1 | K433 | LLHTK **MKYE** QLGKM | 0.80 |
| 2 | K346 | KDYVG **LKIP** ASQDL | 0.80 |
| 3 | K68 | RNRYG **GKYD** LHLTQ | 0.67 |
| 4 | K675 | ERGLG **PKSD** GELYI | 0.61 |
| 5 | K287 | HDWEN **IKSF** LTDQT | 0.59 |
| 6 | K529 | KILAA **QKAE** KLAEK | 0.50 |
| 7 | K957 | TKCDL **FKQF** SYCLV | 0.50 |
| 8 | K922 | KEMNL **PKTF** WNERI | 0.26 |
| 9 | K481 | LRPRE **EKEF** LIKLS | 0.15 |

➢ **P3 protein [Zucchini yellow mosaic virus]**

**Table 5. motifs position in P3 protein [Zucchini yellow mosaic virus]**

| No. | Pos. | Group | Score |
|---|---|---|---|
| 1 | K32 | GIEVW **IKRD** HEIGK | 0.94 |
| 2 | K149 | CVTVR **LKQF** SIFTE | 0.56 |
| 3 | K39 | RDHEI **GKIF** VILEQ | 0.32 |

- **SOPMA Secondary Structure Prediction:**

  Self-Optimized Prediction Method (SOPM) has been described to improve the success rate in the prediction of the secondary structure of proteins. This method (SOPMA) correctly predicts 69.5% of amino acids for a three-state description of the secondary structure (alpha-helix, beta-sheet and coil) in a whole database containing 126 chains of non-homologous (less than 25% identity) proteins. Joint prediction with SOPMA and a neural networks method (PHD) correctly predicts 82.2% of residues for 74% of co-predicted amino acids.In our case we have selected the Window width-17, Similarity threshold -08 and Number of states- 04.

➢ **P3 protein [Watermelon mosaic virus]**

10 20 30 40 50 60 70

| | | | | | | |

GEAQQRMKCETALIKSIFKPKRMIQILEDDPYILLMGLISPSILIHMYRMKHFEKGIELWISKEHSVAKI

chhhhhhhhhhhhhhcccthhhhhhhtttceeeeehcctttheeeehcttchhhhhhhhhcttcchhhh

FIIMEQLTRKIAANDLLLEQLDIIAGTSQKLMDVLEDCPQSAHSYRTAKDLLAIYIERRASNNQLIENGF

hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhcchhhhhttc

VDINDQLYVTHEKIYVDRLKQEWHALSWLEKSSITWQLKRFTPHTEQCLTKKVVEESSAYSRNFVSACFM

Eehhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccccchhhhhccccchhhhhhhhhh

NAQSHLKNVRNTFFRKCDQAWTASVRVLVRFIIATLHKCYSDIVYLVNICLIFSLLVQMVSVLQGIVSTA

hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcchhhhhhhhhhhhhhhhhhhhhhhhhhh

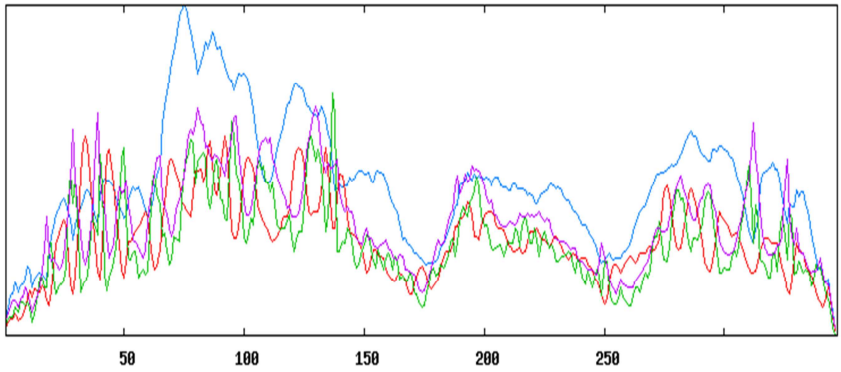
KRDKAFVHMHKRIEDEQAVVHLYEMCEKMENKHPSVEEFLSHVKKVRPELLPVAKSMTGQSEDVSAQ

Hhhhhhhhhhhhhhhhhhhhhhhhhhhhttccccchhhhhhhhhhcchhhhhhhhhhhhtcchhhhc

**Sequence length: 347**

| | | | |
|---|---|---|---|
| Alpha helix | (Hh) | : | 272 is 78.39% |
| $3_{10}$ helix | (Gg) | : | 0 is 0.00% |
| Pi helix | (Ii) | : | 0 is 0.00% |
| Beta bridge | (Bb) | : | 0 is 0.00% |
| Extended strand | (Ee) | : | 11 is 3.17% |
| Beta turn | (Tt) | : | 16 is 4.61% |
| Bend region | (Ss) | : | 0 is 0.00% |
| Random coil | (Cc) | : | 48 is 13.83% |
| Ambigous states | (?) | : | 0 is 0.00% |
| Other states | | : | 0 is 0.00% |

**Fig. 11:**




➢ **Polyprotein, partial [Papaya ring spot virus W]**

10 20 30 40 50 60 70

AAMIESWGYGELTHQIRRFYQWVLEQAPFNELARQGRAPYVSEVGLRRLYTSERGSMDELEAYIDKYFER

hhhhhttchhhhhhhhhhhhhhhtcccchhhhhhttcccchhhhhhhhhhhhhcccccchhhhhhhhhhhhhh

ERGDSPELLVYHESRIADDYQLVCSNNTHVFHQSKNEAVDAGLNEKLKEKEKQKEKEKEKQKEKEKDDAS

ccccccceeeecccccctteeeeettccceeetcccchhhhhhhhhhhcccccccccccccccccccccc

DGNDVSTSTKTGERDRDVNVGTSGTFTVPRIKSFTDKMILPRIKGKTVLNLNHLLQYNPQQIDISNTRAT

ccccccccccccccccccccccceecccchhcccccccttcceeehhheeecccccchhccccc

QSQFEKWYEGVRNDYGLNDNEMQVMLNGLMVWCIENGTSPDISGVWVMMDGETQVDYPIKPLIEHATPSF

hhhhhhhhhhhhhhhcccchhhhhhhhttheeeeecttcccccceeeeecccccccccccchhhtcccchh

RQIMAHFSNAAEAYIAKRNATERYMPRYGIKRNLTDISLARYAFDFYEVNSKTPDRAREAHMQMKAAALR

hhhhhhhhhhhhhhhhhhtcccccchhcchtcchhhhhhhheeeeeecccccchhhhhhhhhhhhhhh

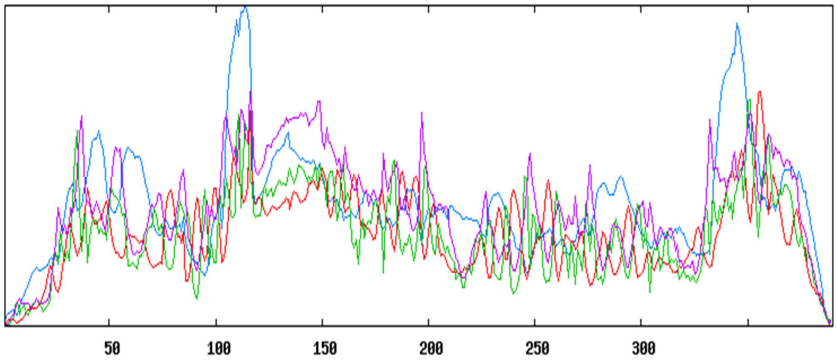
NTSRRMFGMDGSVSNKEENTERHTVEDVNRDMHSLLGMRN

ttcceeeeetccccccccccchhhhhhhhhhhhhhhhcct

**Sequence length: 390**

| | | | |
|---|---|---|---|
| Alpha helix | (Hh) | : | 160 is 41.03% |
| $3_{10}$ helix | (Gg) | : | 0 is 0.00% |
| Pi helix | (Ii) | : | 0 is 0.00% |
| Beta bridge | (Bb) | : | 0 is 0.00% |
| Extended strand | (Ee) | : | 41 is 10.51% |
| Beta turn | (Tt) | : | 23 is 5.90% |
| Bend region | (Ss) | : | 0 is 0.00% |
| Random coil | (Cc) | : | 166 is 42.56% |
| Ambigous states | (?) | : | 0 is 0.00% |
| Other states | | : | 0 is 0.00% |

**Fig. 12:**




➢ **1a protein [Cucumber mosaic virus]**

10 20 30 40 50 60 70

| | | | | | | |

MATSSFNINELVASHGDKGLLATALVDKTAHEQLEEQLQHQRRGRKVYIRNVLGVKDSEVIRNRYGGKYD

hhcchechhhhhhhttcchhhhhhhhhhhhhhhhhhhhhhhhttceeeeehheehcthhhhhhhccccee

LHLTQQEFAPHGLAGALRLCETLDCLDSFPSSGLRQDLVLDFGGSWVTHYLRGHNVHCCSPCLGIRDKMR

eeecccccccchhhhhhhhhhhhhhhhtccctttccccceeeeetcceeeeeetcccccccccccccchhhh

HAERLMNMRKIILNDPQQFDGRQPDFCTQPAADCKVQAHFAISIHGGYDMGFRGLCEAMNAHGTTILKGT

hhhhhhhhhhhhhcccccccccccccceecccccccccheeeeetttccccchhhhhhhhhhttceeeetc

MMFDGAMMFDDQGVIPELNCQWRKIRSAFSETEDVTPLSGKLNSTVFSRVRKFKTMVAFDFINESTMSYV

eeecttheeeettcccchhhhhhhhcccccccccccchhhhhhhhhhhhhhhhtttceeeeeeecctttcceeee

HDWENIKSFLTDQTYSYRGMTYGIERCVIHAGIMTYKIIGVPGMCPPELIRHCIWFPSIKDYVGLKIPAS

cchhhhhhhhhttcccetceeeehhhheehtteeeeeeeecttcccccchheeeccthhhheeecccccc

QDLVEWKTVRILMSTLRETEEIAMRCYNDKKAWMEQFKVILGVLSAKSSTIVINGMSMQSGERIDINDYH

Ccchhhhhhhhhhhhhhhhhhhhhcccccccchhhhhhhhhhhhttcceeeettceectttccchhhhh

YIGFAILLHTKMKYEQLGKMYDMWNASSISKWFAALTRPLRVFLSGVVHALFPTLRPREEKEFLIKLSTF

hhhhhhhhhhhhhhhhhhhhhhhhhhtcccchhhhhhcccchhhhhhhhhhhccchhhhhhhhhhhhhhhccch

VTFNEECSFDGGEEWDVISSAAYVATQAVTDGKILAAQKAEKLAEKLAQPVIEVSDSPEAPSQTPDDTAE

eeehhhhhhcccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhheecccccccccccchhc

VCGKEREVSELDSLSAQTRSPITRVAERATAMLEYAAYEKQLHDTTVSNLKRIWNMAGGDDKRNSLEGNL

ccccccccchhccccccccceecchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhheettcccccccchhhh

KFVFDTYFTVDPMVNIHFSTGRWMRPVPEGVVYSVGYNERGLGPKSDGELYIVNSECVICNSESLSTVTR

hhhhhhhhhhhhhheecccttcceccccteeeeeeecttcccccccceeeeecccccehhhhhhhhhhcc

SLQAPTGTISQVDGVAGCGKTTAIKSIFEPSTDMIVTANKKSAQDVRMALFKSSDSKEACTFVRTADSVL

cccccccceeeeetcccccccceehhhccttceeeeeetcccchhhhhhhhhhccccthhhheehhhhhhe

LNECPTVSRVLVDEVVLLHFGQLCAVMSKLKAVRAICFGDSEQIAFSSRDASFDMRFSKIIPDETSDADT

etcccthhheehhhheecchhhhhhhhhhhhhheeeeecccccceeeeccccceeeeeeeecccccccccc

TFRSPQDVVPLVRLMATKALPKGTRSKYTKWVSQSKVKRSVTSRAIVSVTLVDLDPSRFYITMTQADKAS

ccccchhhhhhhhhhhtcccttcccceeeeecchhhhhhccccceeeeeecccccteeeeeccccchhh

LISRAKEMNLPKTFWNERIKTVHESQGISEDHVTLVRLKSTKCDLFKQFSYCLVALTRHKVTFRYEYCGV

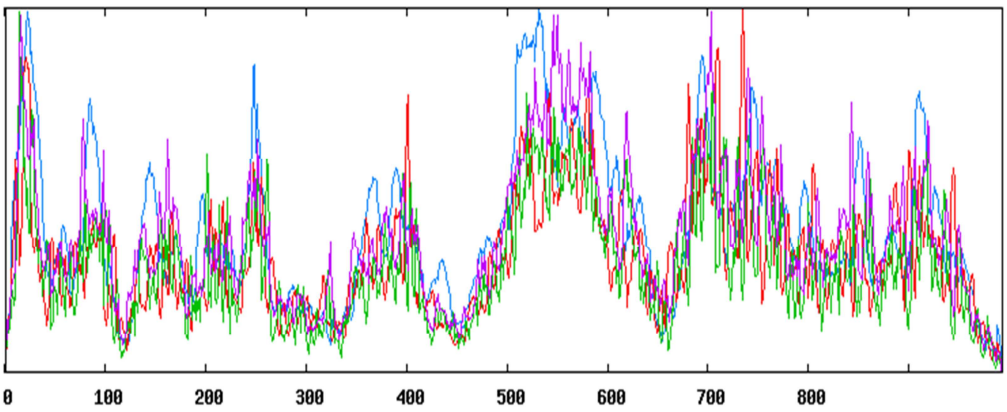
hhhhhhhhtcccchhhhhhhhhhhhhtttccccheeeeeecccccchhhhhhhhheeeehtttcceeeeeccc

LNGDLIAECVARAccccceehhhhhhh

**Sequence length:993**

| | | | |
|---|---|---|---|
| Alpha helix | (Hh) | : | 437 is 44.01% |
| $3_{10}$ helix | (Gg) | : | 0 is 0.00% |
| Pi helix | (Ii) | : | 0 is 0.00% |
| Beta bridge | (Bb) | : | 0 is 0.00% |
| Extended strand | (Ee) | : | 178 is 17.93% |
| Beta turn | (Tt) | : | 66 is 6.65% |
| Bend region | (Ss) | : | 0 is 0.00% |
| Random coil | (Cc) | : | 312 is 31.42% |
| Ambigous states | (?) | : | 0 is 0.00% |
| Other states | | : | 0 is 0.00% |

**Fig. 13:**



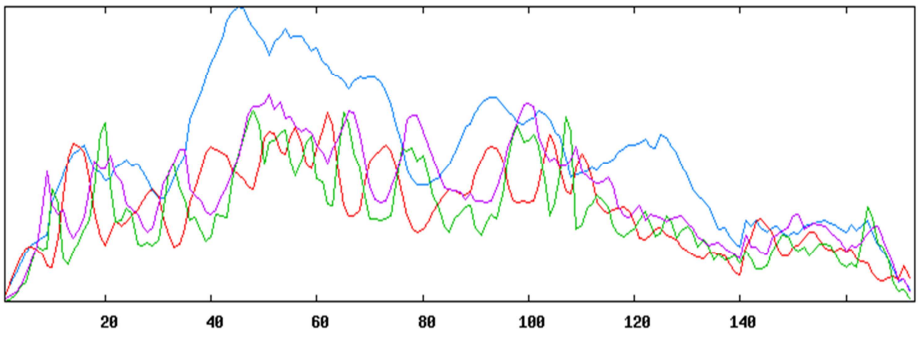
➢ **P3 protein [Zucchini yellow mosaic virus]**

```
    10        20        30        40        50        60        70
     |        |         |         |         |         |         |
PYILLLGMISPTILVHMYRMRHFERGIEVWIKRDHEIGKIFVILEQLTRKVALAEVLVDQLNLISEASPH
cheeeehhcchheeehhcttchhhhhhheeccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
LLEIMKGCQDNQRAYVPALDLLTIQVEREFSNKELKTNGYPDLQQTLFDMWEKMYAKQLHNSWQELSLLE
hhhhhhhccccccchhhhhhhhhhhhhhcccchhhhhttceehhhhhhhhhhhhhhhhhhhhhhhhhhh
KSCVTVRLKQFSIFTERNLIQRAEEGKRASSLQ
hhhheehhccchhhhhhhhhhhhttcccceee
```

**Sequence length: 173**

| | | | |
|---|---|---|---|
| Alpha helix | (Hh) | : | 120 is 69.36% |
| $3_{10}$ helix | (Gg) | : | 0 is 0.00% |
| Pi helix | (Ii) | : | 0 is 0.00% |
| Beta bridge | (Bb) | : | 0 is 0.00% |
| Extended strand | (Ee) | : | 16 is 9.25% |
| Beta turn | (Tt) | : | 7 is 4.05% |
| Bend region | (Ss) | : | 0 is 0.00% |
| Random coil | (Cc) | : | 30 is 17.34% |
| Ambigous states | (?) | : | 0 is 0.00% |
| Other states | | : | 0 is 0.00% |

**Fig. 14:**




- **Predict transmembrane:**

From the analysis of TMHMM result it could be inferred that there are no transmembrane helices present in this sequence so, it is not likely to be a transmembrane protein **.**It also indicates that any of the predicted transmembrane helix in the N-term is not a signal peptide.

➢ **P3 protein [Watermelon mosaic virus]**

Sequence Length : 347

# Sequence Number of predicted TMHs : 1

# Sequence Exp number of AAs in TMHs : 30.61454

# Sequence Exp number, first 60 AAs : 10.07098

# Sequence Total prob of N-in : 0.74772

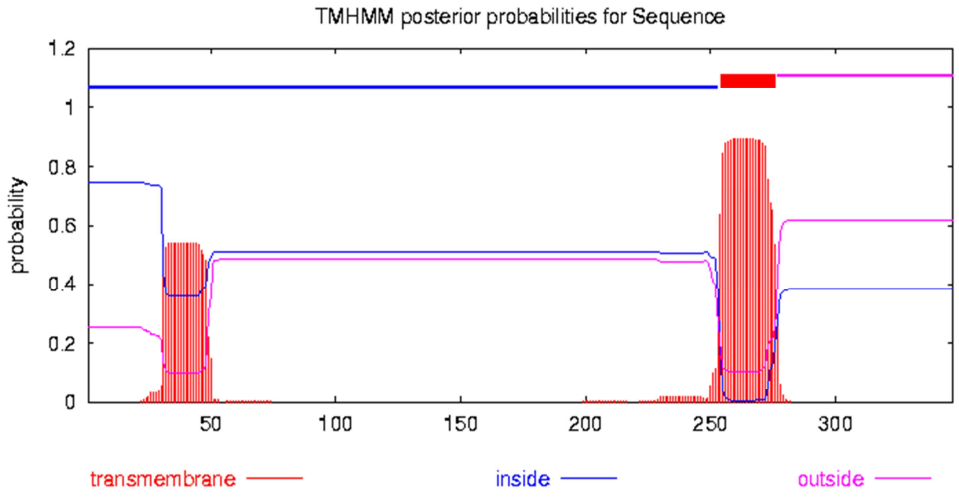
**Sequence POSSIBLE N-term signal sequence**

Sequence TMHMM2.0 inside 1 253

Sequence TMHMM2.0 TMhelix 254 276

Sequence TMHMM2.0 outside 277 347

**Fig. 15:**




➢ **Polyprotein, partial [Papaya ring spot virus W]**

Sequence Length : 390

# Sequence Number of predicted TMHs : 0

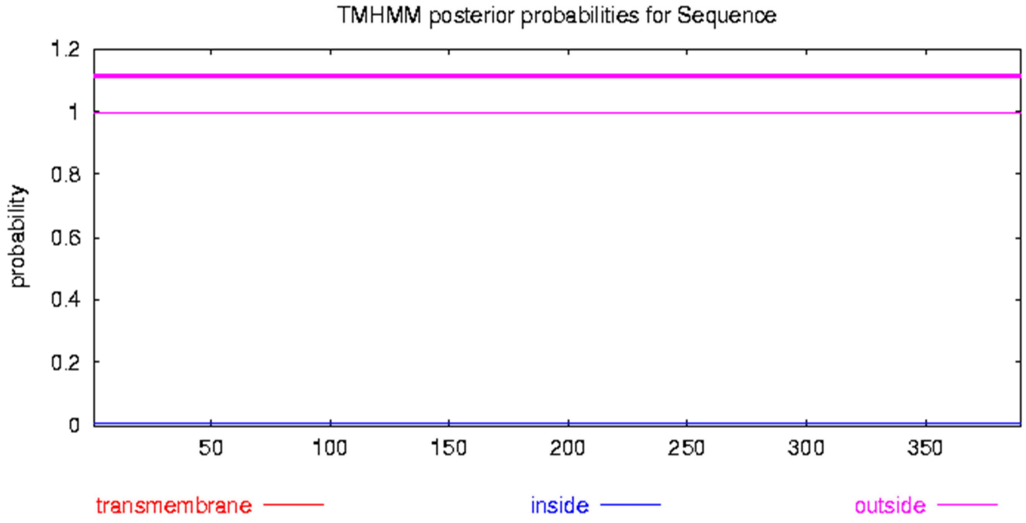
# Sequence Exp number of AAs in TMHs : 0.00127

# Sequence Exp number, first 60 AAs : 0

# Sequence Total prob of N-in : 0.00659

Sequence TMHMM2.0 outside 1 390

**Fig. 16:**



- **1a protein [Cucumber mosaic virus]**

Sequence Length: 993

# Sequence Number of predicted TMHs : 0

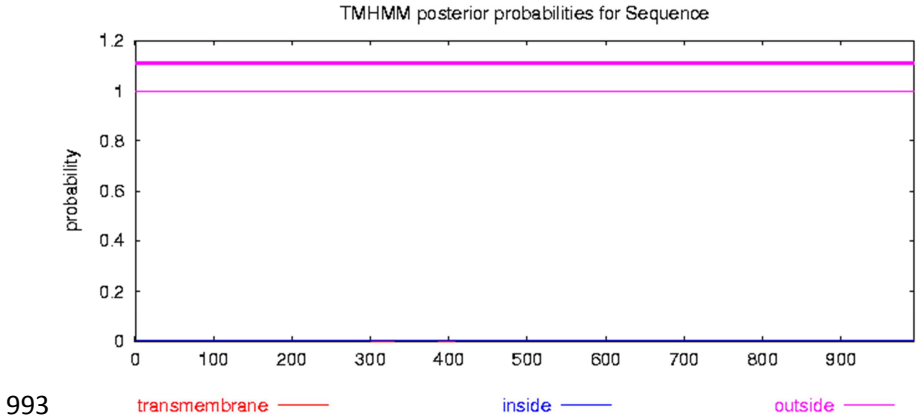
# Sequence Exp number of AAs in TMHs : 0.03494

# Sequence Exp number, first 60 AAs : 0

# Sequence Total prob of N-in : 0.00152

Sequence TMHMM2.0 outside 1

**Fig. 17:**




- **P3 protein [Zucchini yellow mosaic virus]**

Sequence Length : 173

# Sequence Number of predicted TMHs : 0

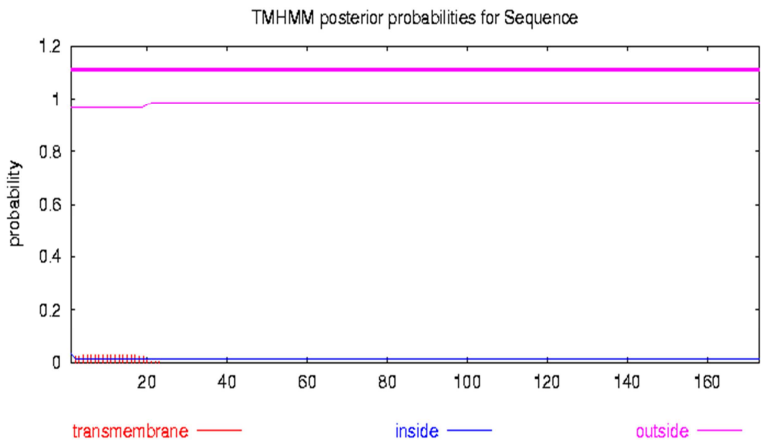
# Sequence Exp number of AAs in TMHs : 0.4565

# Sequence Exp number, first 60 AAs : 0.4565

# Sequence Total prob of N-in : 0.03243

Sequence TMHMM2.0 outside 1 173

**Fig. 18:**



## CONCLUSION

Several viruses that affect cucumbers, melons, pumpkins, squash, and other members of the cucurbit family e.g. Cucumber mosaic virus (CMV), zucchini yellow mosaic virus (ZYMV), watermelon mosaic virus (WMV) and papaya ring spot-W (PRSV-W). All are transmitted from diseased plants to healthy plants by aphids from plant to plant in a non-persistent manner. This means they acquire the virus from an infected plant almost immediately but are able to infect healthy plants for only a short time, usually a few days to a week. Only a small number of aphids are needed to spread the virus throughout the field. CMV is also spread by spotted and striped cucumber beetles.Cucumber mosaic, caused by the cucumber mosaic virus, is one of the most widespread and destructive diseases on cucumber and muskmelon. The disease has been known since the early 1900's, and is now found worldwide. The virus can infect cucumber, squash, muskmelon, and numerous other hosts in 34 plant families, including tomato, lima bean, beet, sweet corn, and sweet potato. Most often, actively growing and mature plants are affected. It rarely infects plants in the seedling stage, but will kill them quickly when it does. It causes a decrease in the number and the quality of the fruit.

Here we have done the project about the sequence analysis and prediction of secondary structures of cucurbits. The Fasta format of the target sequences were retrieved from NCBI. The next step for sequence analysis was performed using the Multiple Sequence Alignment Server CLUSTAL W which involves a progressive strategy for aligning pairs of sequence. The CLUSTAL W server was selected for sequence analysis as it exploits the fact that similar sequences are likely to be evolutionary related and it expressed the degree of similarity in the relatively concise format.A lot more information about linear amino acid sequence was known but full understanding of the biological role of these can only be possible if we clearly analyse the secondary structure of protein. As part of secondary structure prediction process, various online servers were used. The full biological roles were understood by analysing the entire possible aspect sand feature with the help of various softwares.

The study gives the insight into engineering the molecules for better study of the enzyme and obtaining the structure molecule for present and future development of the process.

**Acknowledgement**

Author would like to thank to Department of Computational Biology and Bioinformatics, JSBB, SHIATS Allahabad, U.P.-India for supporting this work by providing a good research environment and related facilities.

## REFERENCES

1. Aramburu, J., Galipienso, L., Lopez, C., Reappearance of Cucumber mosaic virus isolates belonging to subgroup IB in tomato plants in North-eastern Spain. *Journal of Phytopathology*.,**155:**513–518 (2007).
2. Dijkstra, J. and Khan, J. A., Description of positive-sense, singlestranded RNA viruses. Handbook of Plant Virology / Khan J. A., Dijkstra J. (eds). – New York, USA, 253–388 (2006).
3. Edwardson, J.R. and Christie, R.G., Cucumoviruses. Viruses infecting forage legumes.*Gainesville*.,**1:**143–1214 (1987).
4. Finetti-Sialer, M.M., Cillo, F., Barbarossa, L. and Gallitelli, D., Differentiation of cucumber mosaic virus subgroups by RTPCR-RFLP. *Journal of Plant Pathology*.,**81:**145–149 (1999).
5. Francki, R.I.B., Mossop, D.W. and Hatta, T., Cucumber mosaic virus. *CMI/AAB Descriptions of plant viruses*. **213:**1–6 (1979).
6. Hord, M.J., Garcia, A. and Villalobos, H. *et al*., Field survey of Cucumber mosaic virus subgroups I and II in crop plants in Costa Rica. *Plant Disease*.,**85(9):**952–954 (2001).
7. Kaper, J.M and Waterworth, H.E., Cucumoviruses. Handbook of plants virus infections and comparative diagnosis / Kurstak E. (ed.). – Amsterdam, Netherlands, 257–332 (1981).
8. Lin, H.X., Rubio, L. andSmythe, A. et al., Genetic diversity and biological variation among California isolates of Cucumber mosaic virus. *Journal of General Virology*.,**84:** 249–258 (2003).

9. Palukaitis, P., Roossinck, M.J., Diecgen, R.G., Francki, R.I.B., Cucumber mosaic virus. *Advances in Virus Research.*,**41:**280–348 (1992).
10. Roossinck, M.J. and White, P.S.,Cucumovirus isolation and RNA extraction. Plant Virology Protocols from virus isolation to transgenic resistance. Methods in molecular biology / Forster G. D., Taylor S. C. (eds). – Totowa, USA, 189–196 (1998).
11. Roossinck, M.J., Zhang, L. and Hellwald, K.H., Rearrangements in the 5′ nontranslated region and phylogenetic analyses of Cucumber mosaic virus RNA3 indicate radial evolution of three subgroups. *Journal of General Virology*., **73(8):**6752–6758 (1999).
12. Samuitienė, M. and Navalinskienė, M., Occurrence of Cucumber mosaic cucumovirus on ornamental plants in Lithuania. *Žemdirbystė=Agriculture*.,**95(3):**135–143 (2008).
13. Samuitienė, M., Savičienė, A., Virus ogurečnojmozajkinačernojsmorodine v Litve. Zaščitasel'skohozjajstvennyhrastenij v uslovijahprimenenijaintensivnyhtehnologij. –, č. **2:**126 (in Russian) (1987).
14. Staniulis,J., Ankštiniųaugalųvirusiniųirgeltostipoligųsukėlėjai Lietuvoje: gamtosmokslųhabilitacinisdarbas. – Vilnius,. – 75 p. (in Lithuanian) (1994).
15. Yordanova, A. andHristova, D.,Serogroup differentiation of Bulgarian isolates of cucumber mosaic virus. Comptesrendus de l'Academiebulgare des Sciences.,**55(2):**75–80 (2002).
16. Zitikaitė, I., Diversity of vegetable viruses identified in Lithuania.Problemysochranenijabiologičeskogoraznoobrazija i ispol'zovanijabiologičeskihresursov. – Minsk, **1:** 117–119 (2009).