IN SILICO EPITOPE PREDICTION OF COAT PROTEIN GENE OF SUGARCANE MOSAIC VIRUS

AMAL MAHMOUD[1, 2] AND ALI H. HAMED[3]

[1] *Department of Bioinformatics, Genetic Engineering and Biotechnology Research Institute (GEBRI), University of Sadat City, Egypt.*

[2]*Department of Biology, College of Science, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, 31441, Dammam, Saudi Arabia*

[3]*Virus and Phytoplasma Research, Plant Pathology Institute, Agriculture Research Center, Egypt*

Sugarcane mosaic virus (ScMV) is one of the economically important viruses causing losses in various sugarcane-growing regions including Egypt, Natal (South Africa), Hawaii, Argentina, Cuba, Australia, and the USA (Koike and Gillaspie, 1989). Also, the virus infects maize and sorghum (Teakle *et al.,* 1989). In Egypt, ScMV has been recorded in maize and sugarcane (Abd El Fattah *et al*., 2005) and El-Morsi *et al*., 2003).

Sugarcane mosaic virus belongs to the genus Potyvirus, and family Potyviridae (Wu *et al*., 2013 ; Zhu *et al*., 2014). Potyviruses have a single-stranded positive-sense RNA genome. Genomic RNA is translated into two polyproteins that require proteolytic processing to produce ten mature and one fusion protein essential for replication and movement: P1 (translation, modulator of replication), helper component proteinase HC-Pro (silencing suppression and aphid transmission), P3 (virus replication and movement), P3N-PIPO (cell-to-cell movement), 6K1 (formation of replication vesicles), cytoplasmic inclusion protein (CI, helicase involved in virus movement and replication), 6K2 (formation of replication vesicles), genome-linked protein VPg (translation, virus movement, and replication), NIa-Pro (polyprotein processing), NIb (RNA-dependent RNA polymerase), and CP (virus movement, virion formation and aphid transmission) (Nigam *et al*., 2019).

Recently, genomic characterization of ScMV was investigated (Goncalves *et al*., (2011) and Wu *et al*., (2012). The virus is easy to mutate. Therefore, ScMV exists as numerous strains and replicates as complex and dynamic mutant swarms (Goncalves *al*., 2011, Li *et al*., 2013 and Padhi and Ramu 2011). ScMV isolates were classified into five groups (sugarcane, maize, Thailand, Noble sugarcane and Brazil groups) based on the coat protein (CP) coding gene (Xu *al*., 2008). Most of the codons of the CP coding gene

Egypt. J. Genet. Cytol., 48:235-243, July, 2019
Web Site (*www.esg.net.eg*)

proved to be under negative selection and recombinations also existed within the CP cistron (Li *et al*., 2013).

In this study, a ScMV isolate from Upper Egypt was collected in 2015 and tested by RT-PCR. The partial CP coding gene of ScMV was sequenced and compared to those available in GenBank databases. Bioinformatics tools were used to predict essential epitopes of ScMV coat protein for virus diagnosis facilitation.

MATERIALS AND METHODS

ScMV detection using RT-PCR

Sugarcane samples revealed ScMV symptoms were collected from Qena governorate-Egypt in 2015. The total RNA was extracted from sugarcane samples that suspected to be infected with ScMV using Gene Jet™ plant RNA purification kit (Thermo Scientific, USA) according to manufacturer's manual. The total RNA was used as a template for one-step RT-PCR using Verso™ 1-step RT-PCR Kit as the following: 12.8 µl of nuclease–free water, 10 ng/µl of total RNA, 10 µl of 5x buffer, 2 µl of 10 mM dNTPs mix, 10 µl of 5x Q-solution, 3 µl of 1 µM of each primer (ScMV-F :5`-ACACAAGAGCAACCAGAGAGG-3`and ScMV-R: 5`-AGTCAAAGGCATACCGCGCTA-3`), 2 µl of Verso Enzyme Mix, and 0.2 µl of RNase inhibitor in total volume 50 µl. The RT-PCR reaction was carried using PCR temperature cycling: 50 °C for 15 min, 95 °C for 2 min, amplification 30 cycles (95°C for 1 min, 60 °C for 1 min, 72 °C for 1 min) and 72 C° for 10 min**.** The PCR product was cloned in pGEM-T Easy vector using JM109 *E. coli*. The isolation of the pGEM/ScMV-CP plasmid DNA construct was performed using QIAprep Spin Miniprep kit from Qiagen.

DNA sequencing

The ScMV-CP/pGEM clones validated by restriction enzyme digestion were selected for DNA sequencing. Automated DNA sequencing was performed using dRhodamine terminator cycle sequencing ready reaction kit from PE-Applied Biosystems in conjunction with the ABI PRISMTM 310 Genetic Analyzer from PERKIN-ELMER (Applied Biosystems Division, Foster City 6, CA, USA). The forward and the reverse universal sequencing primers (pUC/M13) specific for the pGEM cloning system were used for DNA sequencing.

Sequence analysis

Sequence analysis was performed using BLAST programs from National Center for Biotechnology Information (NCBI), USA, (http://www.ncbi.nlm.nih.gov/Blast).
DNA sequences of ScMV-CP isolates were retrieved from GenBank and were used for performing comparisons against our ScMV-CP isolate. Multiple sequence alignment was performed using Mega 6.0 software (Tamura *et al*., 2013).

MFOLD version 3.2 program was used to predict the effect of mutations on ScMV-CP secondary structure (Zuker, 1989).

The B cell epitopes of ScMV-CP sequence were predicted using Immune epitope databases (IEDB) analysis resource (http://tools.immuneepitope.org/bcell/). The coat protein sequence of ScMV isolates was retrieved from the GenBank database. Using IEDB, multiple sequence amino acid alignment of ScMV-CP was performed to recognize the immunologically relevant regions including antigenicity, solvent accessible regions, and exposed surface area.

RESULTS AND DISCUSSION

Nucleotide and amino acid sequence analysis

Potyviridae is the largest and most economically important family of plant viruses. The advent of modern techniques to determine the antigenic regions of the viral amino acid sequences will help to resolve difficulties in detecting some members of the family. In this study, a partial sequence of a CP ScMV Egyptian isolate was used to detect universal epitopes to detect ScMV.

PCR amplification of the partial CP coding gene of a ScMV Egyptian isolate (ScMV-Egy) isolate resulted in 360bp DNA fragment (Fig. 1). The PCR fragment was sequenced. The DNA sequence was edited and deposited into GenBank (accession numbers: KY436736 and ASK38697 for nucleotide and amino acid sequences, respectively). We compared the DNA and protein sequences of the ScMV-Egyptian isolate coat protein to their homologs in the GenBank (Fig. 2 A and B, respectively). Using BLAST tool, ScMV-Egy showed 97.5–99.2% and 94.9-97.5% for nucleotide and amino acid sequence similarity, respectively). The highest similarity was detected with isolates from Iran and Brazil for nucleotide and amino acid sequence, respectively. According to Moradi, *et al*., (2017), the Iranian isolates shared nucleotide identities of 75.5–99.9% with those of other ScMV isolates available in GenBank, the highest with the Egyptian isolate EGY7-1 (97.5–99.9%). DNA sequence comparisons showed three nucleotide substitutions ($C_{8822}\rightarrow T$, $G_{8831}\rightarrow T$, and $T_{8846}\rightarrow C$) that resulted in an amino acid substitution ($K_{2894}\rightarrow N$). In comparison toother ScMV genes, the DNA sequence identity of the CP coding gene was the highest (Xie *et al*., 2016).

RNA secondary structure prediction of ScMV coat protein

One of deleterious effects of large-scale synonymous substitutions in viral genes is altering the viral RNA secondary structure (Mochizuki, *et al*., 2018). The effect of the three nucleotide substitutions $C_{8822}\rightarrow T$, $G_{8831}\rightarrow T$, and $T_{8846}\rightarrow C$ on RNA secondary structure of ScMV-CP was analyzed using the mFOLD program.

None of the three substitutions resulted in RNA secondary structure change (Fig. 3). El-Absawy *et al*. (2012) reported that one nucleotide substitution in the coat protein of potato virus Y, resulted in RNA secondary structure alteration.

Prediction of antigenic epitopes of the ScMV- CP

Using the multiple amino acid sequence of ScMV-CP (Fig. 2B), the highly conserved region in the coat protein was selected and tested using Immune epitope databases (IEDB) analysis resource. The highest scores obtained with B cell epitopes were TRATREEFDRW2878, DDTQMTVVMSGL2899, EEFDRWYEAI2883, TRATREE2878, and TVVMSGLMVWCI2904 (Table 1). By analyzing the antigenicity, hydrophobicity, solvent accessibility, and exposed surface area, Alam *et al*., (2013) found the potential epitope locations at the sequences 181-MPRYGLVRN-189, 41-THTVPRIKAI-50, and 94 YEAVQLAYDIGETEM-108, for eliciting immune response and targeting for potato virus detection. We previously used *in silico* prediction to predict specific epitopes against bovine ephemeral fever virus (Mahmoud, 2012) and for producing specific vaccine against leishmaniasis in experimental mice (El-Nahas *et al*., 2015).

We suggest that the highly score-epitopes could be used to produce a specific diagnostic tool against ScMV isolates in Egypt and worldwide.

SUMMARY

Sugarcane mosaic virus (ScMV) cause economic losses in the production of sugarcane, maize, sorghum, and some other graminaceous species. Detection of ScMV isolates using a very specific and efficient tool is needed for virus diagnosis and management. In this study, sequence analysis of ScMV partial coat protein (CP) was used to determine conserved sequences-potential epitope regions. PCR amplification of ScMV partial CP gene of an Egyptian isolate of ScMV (ScMV-Egy) resulted in 360 bp. Using multiple sequence analysis, 3 major substitutions were detected, (C_{8822}→T, G_{8831}→T, and T_{8846}→C) and resulted in one amino acid substitution (K_{2894}→N). None of the 3 substitutions affected the predicted RNA structure. Analysis of antigenicity, hydrophobicity, solvent accessibility, and exposed surface area resulted in 5 highly scored epitopes: TRATREEFDRW2878, DDTQMTVVMSGL2899, EEFDRWYEAI2883 TRATREE2878 and TVVMSGLMVWCI2904. The highly score-epitopes should be useful for producing sero-diagnostic ScMV antiserum. Further *in vivo* experiments to validate the predicted epitopes are needed.

ACKNOWLEDGEMENT

The authors thank Dr. Ahmed K. El-Attar, associate professor of molecular virology at the Agricultural Research Centre, Giza

Egypt for his support and help to accomplish this work.

REFERENCES

Abd El Fattah, A. I.; Hanan A. Nour El-Din; Abodoah1, A. M. and Sadik, A. S. (2005). Occurrence of two *Sugarcane mosaic potyvirus* strains in sugarcane. Pak. J. Biotechnol., 2: 1-12.

Allam, M. d. J. ; Ashraf, K. U. M. ; Gupta, S. D. and Emon, M. A. E. K. (2013). Computational approach for the prediction of potential MHC binding peptides and epitope mapping in order to develop serodiagnostic immunogen against *Potato virus Y.* International Journal of Computational Bioinformatics and In Silico Modeling, 2: 186-198.

El-Absawy, E. A ,Khidr, Y. A. Mahmoud, A., Hasan, M. E. and Hemeida, A. A. (2012). Molecular variation of Potato virus Y isolated from Egypt. International J. Virol., 8: 81-89.

El-Morsi, A. A.; Sadik, A. S.; Soweha H. A. and El-Dahlob, S.M. (2003). Partial characterization of an isolate of *Maize dwarf mosaic potyvirus* and production of polyclonal antibodies for virus detection. Annals Agric. Sci. Fac. Agric., Ain Shams Univ., 48: 69-84.

El-Nahas, N. S., Sadek, G. S., Mahmoud, A., Rady, A. A., Oshiba, S. F. and Barakat, R. M. (2015). Immunological, histopathological and immunohistochemical responses to a new B cell epitope-based vaccine against leishmaniasis in experimental mice. Journal of American Science, 11: 1-15.

Goncalves, M. C. ;Galdeano, D. M. ; Maia, I. D. and Chagas, C. M. (2011). Genetic variability of *Sugarcane mosaic virus* causing maize mosaic in Brazil. Pesqui Agropecu Bras., 4: 362-369.

Koike, H & Gillaspie, AG Jr (1989). Mosaic', In Ricaud, C. Egan, BT Gillaspie AG Jr and Hughes, CG (eds.). Diseases of Sugarcane New York, USA: Elsevier, p. 301-322.

Li, Y. Q. ; Liu, R. Y. ; Zhou, T. and Fan, Z. F. (2013). Genetic diversity and population structure of *Sugarcane mosaic virus*. Virus Res., 1: 242–246.

Mochizuki, T. ; Ohara, R.; Roossinck, M. J. (2018). Large-Scale Synonymous Substitutions in Cucumber Mosaic Virus RNA 3 Facilitate Amino Acid Mutations in the Coat Protein. J. of Virol., 92: 1-13.

Padhi, A & Ramu, K (2011). Genomic evidence of intraspecific recombination in *Sugarcane mosaic virus*. Virus Genes, 2: 282-285.

Mahmoud, A (2012). Comparative se-

quence analysis and epitope prediction of the partial glycoprotein G gene of bovine ephemeral fever virus isolate from Egypt**.** Egyptian J. Virol., 9: 49-58.

Moradi, Z.; Nazifi, E. and Mehrvar M. (2017). Occurrence and Evolutionary Analysis of Coat Protein Gene Sequences of Iranian Isolates of Sugarcane mosaic virus. Plant Pathol. J., 33: 296-306.

Nigam, D. ; LaTourrette, K. ; Souza, P. F. N.; and Garcia-Ruiz, H. (2019). Genome-Wide Variation in Potyviruses. Front. Plant Sci., 10: 1439.

Tamura K.; Stecher G.; Peterson D.; Filipski A. and Kumar S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution, 30: 2725-2729.

Teakle, D. S.; Shukla D. D. and Ford, R. E (1989). *Sugarcane mosaic virus*. Descriptions of Plant Viruses No. 342.

Wu, L. J. ; Han, Z.P. ; Wang, S. X. ; Wang, X. T. Sun, A. G. and Zu, X. F. (2013). Comparative proteomic analysis of the plant virus interaction in resistant and susceptible ecotypes of maize infected with *Sugarcane mosaic virus*. J. Proteomics, 89: 124-140.

Wu, L. J. ;Zu, X. F. ; Wang, S. X. and Chen, Y. H. (2012). *Sugarcane mosaic virus*—Long history but still a threat to industry. Crop Prot., 42: 74-78.

Xie, X. ; Chen, W. Fu. Q.; Zhang, P. ; An, T. ; Cui, A. (2016). Molecular Variability and Distribution of *Sugarcane mosaic virus* in Shanxi, China. PLoS ONE, 11: 1-12.

Xu, D.-L.; Park, J.-W. ;Mirkov, T. E. ; Zhou, G.-H. (2008). Viruses causing mosaic disease in sugarcane and their genetic diversity in southern China. Arch. Virol., 153: 1031-1039.

Zhu, M. ; Chen, Y. T. ; Ding, X. S. ; Webb, S. L. ; Zhou, T. ; Nelson, R. S. and Fan, Z. F. (2014). Maize Elongin C interacts with the viral genome-linked protein, VPg, of *Sugarcane mosaic virus* and facilitates virus infection. New Phytol., 4: 1291-1304.

Zuker, M (1989). On finding all suboptimal foldings of an RNA molecule. *Science*, 244: 48–52.

Table (1). Prediction of antigenic peptides of ScMV coat protein.

Method	Epitope	Position*	Size
Bepipred Linear Epitope Prediction	TRATREEFDRW	2878-2888	11
Chou & Fasman Beta-Turn Prediction	DDTQMTVVMSGL	2899-2910	12
Emini Surface Accessibility Prediction	EEFDRWYEAI	2892-2883	10
Karplus & Schulz Flexibility Prediction	TRATREE	2878-2884	7
Kolaskar & Tongaonkar Antigenicity	TVVMSGLMVWCI	2904-2915	12

Amino acid positions according to GenBank polyprotein of ScMV-VER1 accession ABW87761.1*

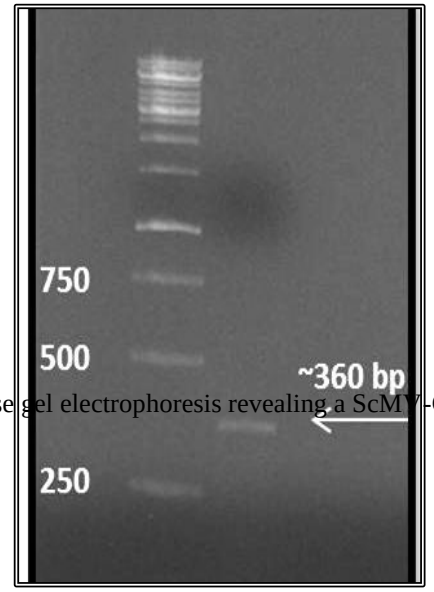
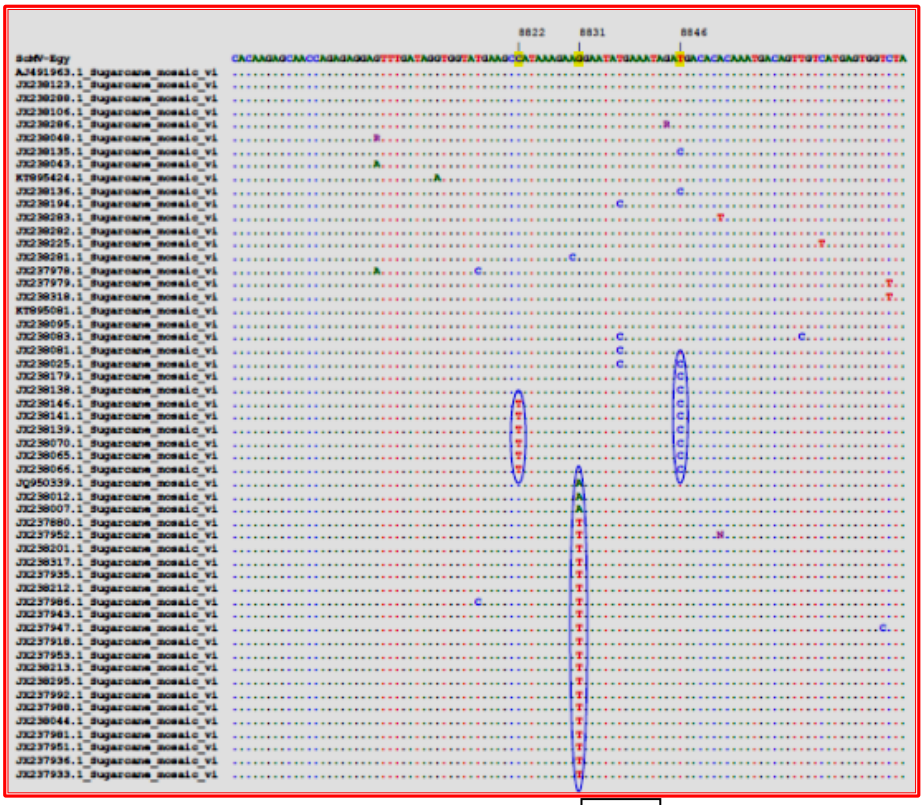
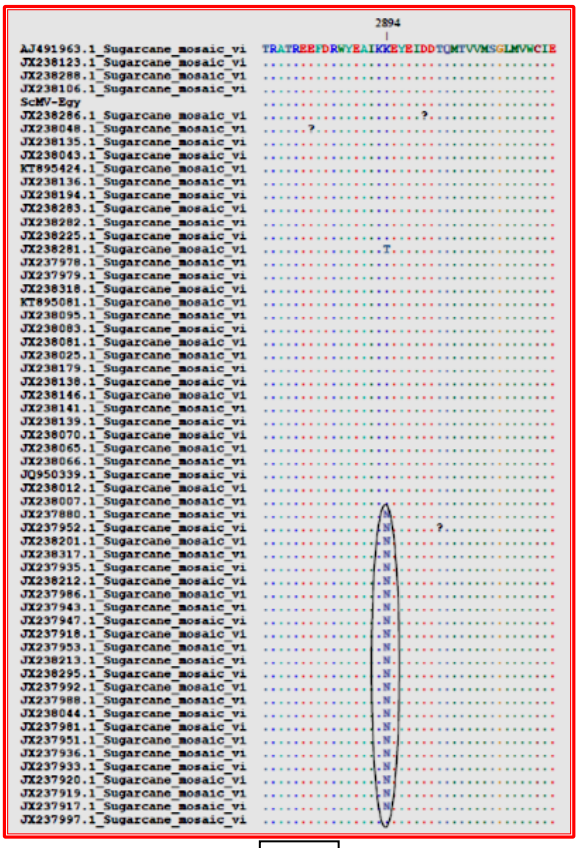


Fig. (1). Agarose gel electrophoresis revealing a ScMV-CP PCR 360 bp fragment.



A



B

Fig. (2). Multiple nucleotide (A) and amino acid (B) sequence alignments of Egyptian ScMV isolate coat protein and their homologs in GenBank database. Nucleotide and amino acid positions correspond to polyprotein of ScMV isolate (accession number EU091075.1 and ABW87761.1, for nucleotide and amino acid sequence, respectively).

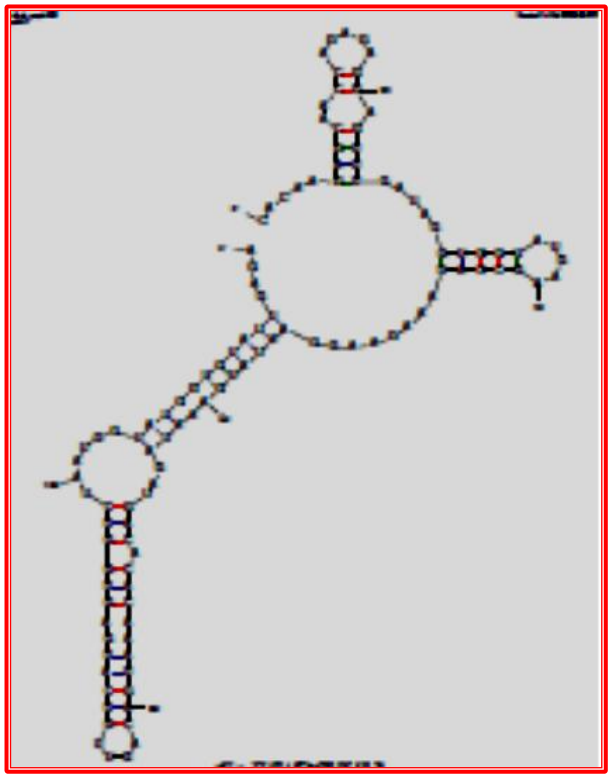


Fig. (3). Prediction of RNA secondary structure of the Egyptian ScMV coat protein