International Journal of Cell Science and Biotechnology
E-ISSN 2320-7574

Available at Available at https://ijcsb.evegenis.org/

Original Research Article

Bio-computational Analysis of fusicoccin from Tomato Plant

Mandeep Singh, Mahesh Pattabhiramaiah* and Swetha R

Centre for Applied Genetics, Jnanabharathi, Bangalore University, Bangalore-560056, India

Accepted 05 Jan 2016, Available online 20 Jan 2016, **Vol.5 (2016)**

Abstract

Fusicoccin(FC) phytotoxin is produced by fungus Fusicoccum amygdali, the protein produced causes the wilt which in turn causes stomatal opening in all green plants. The FC protein sequences were retrieved from NCBI database. Molecular weight of fusicoccin ranged between 25791.8-29431.9 Da. The Aliphatic index of protein ranges from79.80 -92.58, which infers positive factor for thermostability. Instability index predicts that only one FC sequence is stable with 37.97. Primary structure analysis using CLC workbench revealed more of hydrophobic residues. Secondary structure of protein using SOPMA tool predict most of helices ranged from 54.79%-68.67%, least being turns being 4.98%. 3D structure was predicted using Swiss model server. Protein functionality was predicted by SVM method confirms the presence of All DNA binding sites, Metal binding sites etc. Muscle online server was used to reveal conserved and semi-conserved regions. RNA structure was predicted by Genbee online server.

Keywords: *Fusicoccin, NCBI, SOPMA, Genbee, instability index, MUSCLE, Genbee.*

Introduction

Fusicoccin (FC), a diterpene glucoside, is a non-specific phytotoxin produced by the fungus *Fusicoccum amygdali*. Fusicoccin contains three fused carbon rings and another ring which contains an oxygen atom and five carbons. The fusicoccin produces a toxin which is responsible for the wilting of the trees as a result of an irreversible opening of the stomata. Other symptoms of FC which extend beyond the narrow host range of the fungus (Marre, 1979) include the stimulation of cell enlargement, nutrient and ion uptake, seed germination, ethylene production (Malerba *et al*, 1995), and dark CO2 fixation into malate (Brown and Outlaw, 1982). All these FC effects have been attributed to a single cause, i.e., the FC-mediated activation of the plasma membrane H+ -ATPase and, concomitant with the resulting proton extrusion, the hyperpolarization of the plasma membrane, the acidification of the apoplast, and the alkalinization of the cytoplasm (Marre, 1980). FC-mediated activation of the plasma membrane H + -ATPase is not restricted to peach and almond trees, but rather appears to be a general phenomenon in all green plants (Marre, 1979), due to the presence of FC binding protein (FCBP) with high binding affinity for FC (Dohrman *et al,* 1977) The toxin interferes with the endogenous modulation of proton pump activity (Johansson *et al*,1993) by regulatory 14-3-3 proteins (Jahn *et al*, 1997; Oecking *et al*, 1997).

*Corresponding author: **M. M. Abdelrahman**
DOI: https://doi.org/10.14741/ijcsb/v.5.1.1

Plants under attack from pathogens have a range of defense responses at their disposal to fight off the infecting organism. The effect of the fungal toxin fusicoccin (FC) on the tomato (*Lycopersicon esculentum*) is to stimulate a quick acidification of the plant cell wall. FC was further shown to antagonize the system induced membrane depolarization in tomato mesophyll cells (Moyen and Johannes 1996). Fusicoccin increases the susceptibility of a tissue to death by desiccation, and its precise site and mode of action remain obscure. Induces several components of plant pathogen resistance responses, including defense hormone biosynthesis and pathogenesis-related (PR) gene expression. Progressive de-acetylation of fusicoccin reduces the degree of physiological activity as a wilting agent, but tomato shoots showed a novel response, involving an increase of fresh weight, after taking up small amounts of desacetyl-desisopentenyl fusicoccin.

The current pace of high-throughput proteome sequencing programs coupled with high-throughput functional proteomic screens has provided researchers with a bewildering array of sequence and biological data to contend with. Identification of proteins of interest from a particular biological study requires the application of bioinformatics tools to process and prioritize the data. Sequence analysis and physicochemical characterization of proteins using biocomputation tools have been done by many researchers and reported (Ashokan *et al*, 2011; Mahesh and Madhu, 2015; Mahesh *et al*, 2015; Praveen and Mahesh, 2015). From a protein function

standpoint, transfer of annotation from known proteins to a novel target is currently the only practical way to convert vast quantities of raw sequence data into meaningful information. New bioinformatics tools now provide more sophisticated methods to transfer functional annotation, integrating sequence, family profile and structural search methodology. The importance of these approaches to medical research is increasing as we move to annotate the proteome through functional and structural genomic efforts.

The wealth of fusicoccin sequence information that has been made publicly available in recent years requires the development of high-throughput functional genomics and proteomics approaches for its analysis. Such approaches need suitable data integration procedures and a high level of annotation in order to gain maximum benefit from the results generated.

Detailed knowledge of fusicoccin and their properties can be revealed through biological and biochemical properties. The physiochemical and the structural properties of the proteins are well understood and analyzed with the use of computational tools. The statistics about a protein sequence such as number of amino acid, frequency is predicted by CLC work bench (http://w.w.w..clcbio.com/index.php?id=28). Simple Modular Architecture Research Tool (SMART) is a biological database that is used in the identification and analysis of protein domains within protein sequences (Schultz *et al*, 1998 ; Letunic *et al*, 2009). Sequence length, and the physico-chemical properties of proteins such as molecular weight, atomic composition, extinction coefficient, GRAVY, aliphatic index, instability index, etc. can be computed by ProtParam.

The TMpred program makes a prediction of membrane-spanning regions and their orientation. The algorithm is based on the statistical analysis of TMbase, a database of naturally occurring transmembrane proteins. The prediction is made using a combination of several weight-matrices for scoring (Hofmann & Stoffel, 1993). MUSCLE stands for **M**ultiple **S**equence **C**omparison by **L**og-**E**xpectation. MUSCLE is claimed to achieve both better average accuracy and better speed than ClustalW2 or T-Coffee. The protein 3D model and its characteristics can be predicted by Swiss model server (Tsetlin and Hucho, 2004). Protein homology modeling (Garnier and Robson,1978 ; Joyce *et al*, 2004 ; Warren *et al*, 2006) and analogy recognition is made through Phyre2 online server. Reverse Translate accepts a protein sequence as input and uses a codon usage table to generate a DNA sequence representing the most likely non-degenerate coding sequence. A consensus sequence derived from all the possible codons for each amino acid is also returned. Use Reverse Translate when designing PCR primers to anneal to an unsequenced coding sequence from a related species. Further Computer-aided techniques for the efficient identification and optimization of novel molecules with a desired biological activity have become a part of the drug discovery process.

Keeping in view the importance and applications of fusicoccin in tomatoes, computational analysis was performed to determine the physicochemical characteristics of fusicoccin family so as to pave the way to find out better understanding and novel response. The study presents novel insights into the structural, functional, annotational features of fusicoccin in tomatoes.

Materials and methods

Protein sequence retrieval: The Protein Sequences of Fusicoccin (10 sequences) were retrieved in FASTA format from NCBI database (Table1) .

Amino acid Composition: The amino acid composition of selected proteins were computed using the tool CLC free workbench (www.clc.bio.com/..../clc-main-workbench), tabulated in (Table-2).

Primary structure analysis: Counts of hydrophobic and hydrophilic residues were calculated from the primary structure analysis by CLC workbench (Table-3).

Physio-chemical parameters: (http: //web. expasy. org/protparam/). The physicochemical parameters such as theoretical isoelectric point (Ip), molecular weight, total number of positive and negative residues, extinction coefficient, instability index(Gill and Von Hippel, 1989) aliphatic index (Eisenhaber *et al*, 1996) and grand average hydropathy (GRAVY) (Kitchen *et al*, 2007) were computed using the Expasy‘s ProtParam server (Mugilan *et al*, 2010), and tabulated in (Table-4).

Secondary structure prediction: (https://npsaprabi.ibcp.fr/cgibin/npsa_automat.pl?page=npsa_sopma.html). The secondary structure was predicted by self-optimized prediction method with alignment by SOPMA server (Ashokan *et al*, 2011)(Table-5).

Domain architecture analysis: (http://smart.embl-heidelberg.de/) Domain organization and domain composition was analyzed using Simple Modular Architecture Research Tool (SMART) (Table-6).

SVM prot analysis: Which is a protein function prediction tool, and classification of distantly related proteins were analyzed (http://jing.cz3.nus.edu.sg/cgi-bin/svmprot.cgi.) (Table-7).

Trans-membrane region prediction: TMpred was used to predict trans-membrane helices. The TMpred software is available through internet access (http://www.ch.embnet.org/software/TMPRED_form.html)(Fig:1).

Table: 1 Selected Fusicoccin sequences retrieved from NCBI

SR NO	Species	ID	Length	Sequence
1	*Lycopersicum solanum*	NP_ 001296299.1	254aa	MAREENVYMAKLAEQAERYEEMVQFMEKVSTSLGSEELTVEERNLLSVAYK NVIGARRASWRIISSIEQKEESRGNEEHVKCIKEYRSKIESELSDICDGILKLLDS NLIPSASNGDSKVFYLKMKGDYHRYLAEFKTGAERKEAAESTLSAYKAAQDIA NTELAPTHPIRLGLALNFSVFYYEILNSPDRACNLAKQAFDEAIAELDTLGEES YKDSTLIMQLLRDNLTLWTSDMQDDGADEIKETKNDNEQQ
2	*Lycopersicum solanum*	NP_ 001234267.1	261aa	MASSKERESLVYIARLAEQAERYDEMVDAMKNVANLDVELTVEERNLLSVGY KNVVGSRRASWRILSSIEQKEDARGNEQNVKRIQGYRQKVESELTDICNNIMT VIDEHLIPSCTAGESTVFYYKMKGDYYRYLAEFKTGDDKKEVSDLSLKAYQTA TTTAEAELPITHPIRLGLALNFSVFYYEIMNSPERACQLAKQVFDEAISELDSL NEDNYKDGTLILQLLRDNLTLWTSDIPEDGEEAPKGDAANKVGAGEDAE
3	*Lycopersicum solanum*	NP_ 001234097.1	258aa	MASPREENVYMAKLAEQAERYEEMVEFMEKVVAAADGAEELTVEERNLLSV AYKNVIGARRASWRIISSIEKEESRGNEDHVASIKEYRSKIESELTSICNGILKLL DSKLIGSAATGDSKVFYLKMKGDYHRYLAEFKTGAERKEAAENTLSAYKAAQ DIANAELAPTHPIRLGLALNFSVFYYEILNSPDRACNLAKQAFDEAIAELDTLG EESYKDSTLIMQLLRDNLTLWTSDMQDDGTDEIKEATPKPDDNE
4	*Lycopersicum solanum*	NP_ 001234007.1	260aa	MADSSREENVYLAKLAEQAERYEEMIEFMEKVAKTADVEELTVEERNLLSVA YKNVIGARRASWRIISSIEQKEESRGNEDHVNTIKEYRSKIEAELSKICDGILSL LESNLIPSASTAESKVFYLKMKGDYHRYLAEFKTGTERKEAAENTLLAYKSAQ DIALAELAPTHPIRLGLALNFSVFYYEILNSPDRACNLAKQAFDEAISELDTLG EESYKDSTLIMQLLRDNLTLWTSDNADDVGDDIKEASKPESGEGQQ
5	*Lycopersicum solanum*	NP_ 001234677.1	229aa	MEKVSNSLGSEELTVEERNLLSVAYKNVIGARRASWRIISSVEQKEESRGNEE HVNSIREYRSKIENELSKICDGILKLLDSKLIPSATSGDSKVFYLKMKGDYHRYL AEFKTGAERKEAAESTLTGYKAAQDIASAELAPTHPIRLGLALNFSVFYYEILN SPDRACNLAKQAFDEAIAELDTLGEESYKDSTLIMQLLRDNLTLWTSDMQDD GADEIKEDPKPEEKN
6	*Lycopersicum solanum*	NP_ 001234278.1	252aa	MAALIPENLSREQCLYLAKLAEQAERYEEMVQFMDKLVLNSTPAGELTVEER NLLSVAYKNVIGSLRAAWRIVSSIEQKEESRKNEEHVHLVKEYRGKVENELSQ VCAGILKLLESNLVPSATTSESKVFYLKMKGDYYRYLAEFKIGDERKQAAEDT MNSYKAAQEIALTDLPPTHPIRLGLALNFSVFYFEILNSSDKACSMAKQAFEE AIAELDTLGEESYKDSTLIMQLLRDNLTLWTSDAQDQLDES
7	*Lycopersicum solanum*	NP_ 001234637.1	252aa	MEKEREKQVYLARLAEQAERYDEMVEAMKAIAKMDVELTVEERNLVSVGYK NVIGARRASWRILSSIEQKEESKGHEQNVKRIKTYRQRVEDELTKICSDILSVI DEHLVPSSTTGESTVFYYKMKGDYYRYLAEFKAGDDRKEASEQSLKAYEAAT ATASSDLAPTHPIRLGLALNFSVFYYEILNSPERACHLAKQAFDEAIAELDSLS EESYKDSTLIMQLLRDNLTLWTSDLEEGGEHSKGDERQGEN
8	*Lycopersicum solanum*	NP_ 001234272.1	261aa	MASSKERENFVYVAKLAEQAERYDEMVEAMKNVANMDVELTVEERNLLSV GYKNVVGSRRASWRILSSIEKEESRGNEQNVKRIKEYLQKVESELTNICNDIM VVIDQHLIPSCSAGESTVFYHKMKGDYYRYLAEFKAGNDKKEVAELSLKAYQA ATTAAEAELAPTHPIRLGLALNFSVFYYEIMNSPERACHLAKQAFDEAISELDS LNEDSYKDSTLIMQLLRDNLTLWTSDLPEDAEDAQKGDATNKAGGGEDAE
9	*Lycopersicum solanum*	CAA 65148.1	255aa	MASPREENVYMANVADEAERYEEMVEFMEKVVAALNGEELTVEERNLLSVA YKNVIGARRASWRIISSIEQKEESRGNEDHVASIKKYRSQIENELTSICNGILKL LDSKLIGSAATGDSKVFYLKMKGDYYRYLAEFKTGTERKEAAENTLSAYKSAQ DIANGELAPTHPIRLGLALNFSVFYYEILNSPDRACNLAKQAFDEAIAELDTLG EESYKDSTLIMQLLRDNLTLWTSDMQDDGTDEIKEPSKADNE
10	*Lycopersicum solanum*	P9320 6.2|14331_SOLLC	249aa	MALPENLTREQCLYLAKLAEQAERYEEMVKFMDKLVIGSGSSELTVEERNLL SVAYKNVIGSLRAAWRIVSSIEQKEEGRKNDEHVVLVKDYRSKVESELSDVCA GILKILDQYLIPSASAGESKVFYLKMKGDYYRYLAEFKVGNERKEAAEDTMLA YKAAQDIAVAELAPTHPIRLGLALNFSVFYYEILNASEKACSMAKQAFEEAIAE LDTMGEESYKDSTLIMQLLRDNLTLWTSDMQEQMDEA

Table 2: Representation of frequency of amino acids in Fusicoccin

Amino acid	**|NP_001296299.1**	**NP_001234267.1**	**NP_001234097.1**	**|NP_001234007.1**	**NP_001234677.1**	**|NP_001234278.1|**	**NP_001234637.1**	**NP_001234272.1**	**|emb|CAA65148.1**	**|P93206.2|14331_SOLLC**
Alanine (A)	0.094	0.088	0.12	0.104	0.087	0.095	0.091	0.107	0.102	0.108
Cysteine (C)	0.012	0.011	0.008	0.008	0.009	0.012	0.008	0.011	0.008	0.012
Aspartic Acid (D)	0.063	0.073	0.066	0.062	0.066	0.048	0.06	0.061	0.063	0.052

Glutamic Acid (E)	0.126	0.111	0.124	0.131	0.118	0.119	0.131	0.119	0.122	0.12
Phenylalanine (F)	0.024	0.019	0.023	0.023	0.022	0.028	0.02	0.023	0.024	0.024
Glycine (G)	0.039	0.05	0.043	0.038	0.048	0.032	0.044	0.042	0.047	0.04
Histidine (H)	0.012	0.008	0.012	0.012	0.013	0.012	0.02	0.015	0.008	0.008
Isoleucine (I)	0.059	0.054	0.058	0.062	0.061	0.044	0.048	0.042	0.059	0.048
Lysine (K)	0.071	0.061	0.07	0.069	0.079	0.067	0.071	0.069	0.067	0.072
Leucine (L)	0.102	0.096	0.097	0.108	0.109	0.131	0.095	0.092	0.098	0.116
Methionine (M)	0.028	0.023	0.027	0.019	0.017	0.028	0.024	0.031	0.027	0.04
Asparagine (N)	0.051	0.054	0.047	0.046	0.048	0.044	0.028	0.057	0.059	0.032
Proline (P)	0.016	0.023	0.023	0.019	0.026	0.024	0.016	0.019	0.02	0.016
Glutamine (Q)	0.035	0.034	0.023	0.027	0.022	0.044	0.036	0.034	0.024	0.036
Arginine (R)	0.051	0.054	0.05	0.05	0.052	0.044	0.063	0.046	0.051	0.044
Serine (S)	0.087	0.069	0.074	0.088	0.096	0.083	0.087	0.077	0.078	0.076
Threonine (T)	0.043	0.054	0.047	0.046	0.044	0.044	0.048	0.038	0.047	0.032
Valine (V)	0.035	0.061	0.039	0.038	0.035	0.052	0.052	0.061	0.043	0.06
Tryptophan (W)	0.008	0.008	0.008	0.008	0.009	0.008	0.008	0.008	0.008	0.008
Tyrosine (Y)	0.043	0.05	0.043	0.042	0.039	0.044	0.052	0.046	0.047	0.052

Table 3: Hydrophobic and Hydrophilic residues content computed by CLC Workbench

Type	NP_001296299.1	|NP_001234267.1	|NP_001234097.1	|NP_001234007.1	|NP_001234677.1	|NP_001234278.1	|NP_001234637.1	|NP_001234272.1	|emb|CAA65148.1	P93206.2|14331_SOLLC
Hydrophobic (A,F,G,I,L,M,P,V,W)	0.406	0.421	0.438	0.419	0.415	0.44	0.397	0.425	0.427	0.462
Hydrophilic (C,N,Q,S,T,Y)	0.272	0.272	0.24	0.258	0.258	0.27	0.258	0.264	0.263	0.241
Other	0.323	0.307	0.322	0.323	0.328	0.29	0.345	0.31	0.31	0.297

Table 4: Parameters computed by Expasy ProtParam

ID no	**PI**	**Mol wt**	**-R**	**+R**	**EC**	**II**	**AI**	**GRAVY**
NP_ 001296299.1	4.72	28879.3	48	31	27515	46.05	82.68	-0.591
NP_ 001234267.1	4.61	29478.9	48	30	30495	43.40	84.87	-0.531
NP_ 001234097.1	4.68	28967.4	49	31	27515	40.39	83.72	-0.508
NP_ 001234007.1	4.66	29338.8	50	31	25515	43.74	87.54	-0.517
NP_ 001234677.1	4.88	25791.8	42	30	24535	41.60	85.28	-0.588
NP_ 001234278.1	4.80	28624.4	42	28	27515	40.77	92.58	-0.356
NP_ 001234637.1	4.96	28814.2	48	34	30495	48.02	79.80	-0.659
NP_ 001234272.1	4.74	29431.9	47	30	29005	47.55	80.80	-0.538
CAA 65148.1	4.66	28754.1	47	30	29005	40.33	83.88	-0.522
P9320 6.2|14331_SOLLC	4.76	28219.2	43	29	30495	37.97	92.53	-0.269

Table 5: Representation of helix, sheet, turn, coils by through online tool SOPMA

	ID no	NP_001296299.1	NP_001234267.1	NP_001234097.1	NP_001234007.1	NP_001234677.1	NP_001234278.1	NP_001234637.1	NP_001234272.1	CAA 65148.1	P93206.2|14331_SOLLC
Helix(H)	Residue totals	173	143	169	176	148	168	152	160	159	171
	Percentage%	68.11	54.79	65.50	67.69	64.63	66.67	60.32	61.30	62.35	68.67
Sheet(E)	Residue totals	25	38	27	22	26	30	35	34	27	34
	Percentage%	9.84	14.56	10.47	8.46	11.35	11.90	13.89	13.03	10.59	13.65
Turn(T)	Residue totals	10	13	8	9	11	9	9	11	12	7
	Percentage%	3.94	4.98	3.10	3.46	4.80	3.57	3.57	4.21	4.71	2.81
Coils(C)	Residue totals	46	67	54	53	44	45	56	56	57	37
	Percentage%	18.11	25.67	20.93	20.38	19.21	17.86	22.22	21.46	22.35	14.86

Table 6: SMART analysis of Fusicoccin

ID no	Start	End	E-value
NP_001296299.1	11	254	1.24e-202
NP_001234267.1	7	248	7.12e-168
NP_001234097.1	5	249	2.55e-204
NP_001234007.1	6	249	2.63e-206
NP_001234677.1	1	221	5.63e-174
NP_001234278.1	11	251	2.36e-176
NP_001234637.1	5	246	2.93e-172
NP_001234272.1	7	248	3.61e-172
CAA 65148.1	5	248	4.84e-196
P9320 6.2|14331_SOLLC	9	248	4.95e-170

Table 7: SVM prot analysis of Fusicoccin

Protein family name								
Accession number	All DNA binding		Metal binding		Zinc binding		Other protein families	
	R value	P value	R value	P value	R value	P value	R value	P value
NP_001296299.1	**1.6**	**76.2**	**1**	**58.6**	**NA**	**NA**	**Repressor 1**	**58.6**
NP_001234267.1	1.7	78.4	1	58.6	NA	NA	NA	NA
NP_001234097.1	1	58.6	1	58.6	NA	NA	NA	NA
NP_001234007.1	1.4	71.3	NA	NA	NA	NA	NA	NA
NP_001234677.1	1	58.6	NA	NA	NA	NA	(MTB) 1	58.6
NP_001234278.1	1.6	76.2	1.1	62.6	NA	NA	NA	NA
NP_001234637.1	1.4	71.3	NA	NA	1.5	73.8	Mg binding 1	58.6
NP_001234272.1	1.4	71.3	NA	NA	1.5	73.8	Mg binding 1	58.6
CAA 65148.1	1.4	71.3	NA	NA	NA	NA	NA	NA
P9320 6.2|14331_SOLLC	NA	NA	1	58.6	NA	NA	DNA repair 1	58.6

(R value and P values are in percentage)

Table 8: RNA structure stems with free energy (ref CAA 65148.1)

Stem no	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Free energy(Kkal/mol)	-20.4000000	-15.9000000	-15.0000000	-12.7000000	-12.7000000	-11.7000000	-11.7000000	-10.9000000	-10.7000000	-9.6000000	-9.6000000	-9.4000000	-9.2000000	-9.0000000	-9.0000000	-8.9000000	-8.3000000	-8.2000000	-8.1000000	-8.0000000	-7.7000000	-7.3000000	-7.3000000	-7.0000000	-7.0000000	-6.7000000	-6.6000000

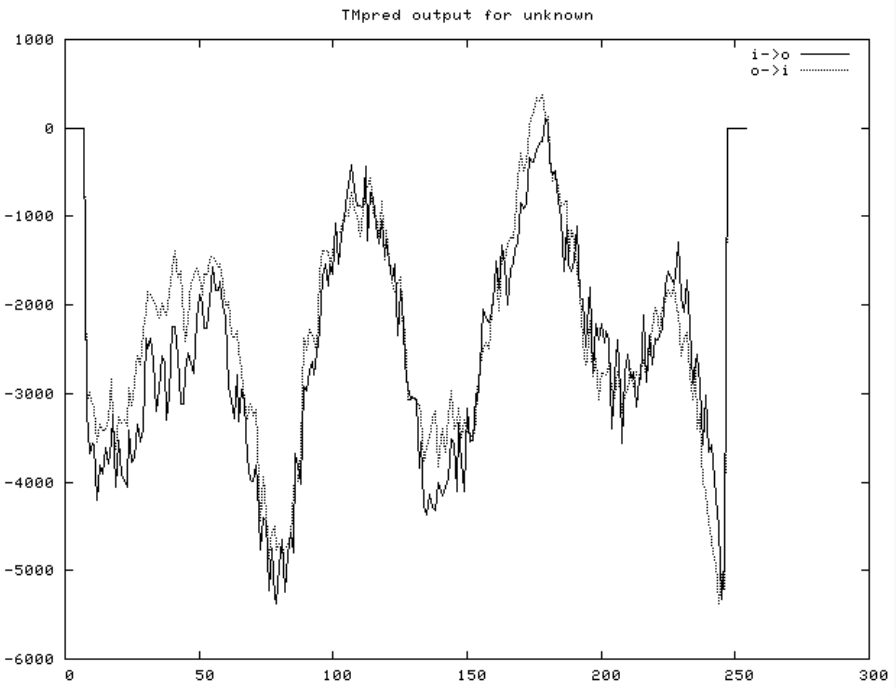


Fig.1 TMPred graph: (ref CAA 65148.1)

```
gi|350539221|ref|NP_001234637.1|      SSTTGESTVFYYKMKGDYYRYLAEFKAGDDRKEASEQSLKAYEAATATASSDLAPTHPIR
gi|350539761|ref|NP_001234267.1|      SCTAGESTVFYYKMKGDYYRYLAEFKTGDDKKEVSDLSLKAYQTATTTAEAELPITHPIR
gi|350536935|ref|NP_001234272.1|      SCSAGESTVFYHKMKGDYYRYLAEFKAGNDKKEVAELSLKAYQAATTAAEAELAPTHPIR
gi|350536755|ref|NP_001234007.1|      SASTAESKVFYLKMKGDYHRYLAEFKTGTERKEAAENTLLAYKSAQDIALAELAPTHPIR
gi|350538649|ref|NP_001234097.1|      SAATGDSKVFYLKMKGDYHRYLAEFKTGAERKEAAENTLSAYKAAQDIANAELAPTHPIR
gi|1771172|emb|CAA65148.1|            SAATGDSKVFYLKMKGDYYRYLAEFKTGTERKEAAENTLSAYKSAQDIANGELAPTHPIR
gi|350534900|ref|NP_001234677.1|      SATSGDSKVFYLKMKGDYHRYLAEFKTGAERKEAAESTLTGYKAAQDIASAELAPTHPIR
gi|823683788|ref|NP_001296299.1|      SASNGDSKVFYLKMKGDYHRYLAEFKTGAERKEAAESTLSAYKAAQDIANTELAPTHPIR
gi|350539807|ref|NP_001234278.1|      SATTSESKVFYLKMKGDYYRYLAEFKIGDERKQAAEDTMNSYKAAQEIALTDLPPTHPIR
gi|26454605|sp|P93206.2|14331_SO       SASAGESKVFYLKMKGDYYRYLAEFKVGNERKEAAEDTMLAYKAAQDIAVAELAPTHPIR
                                      *.: .:*.*** ******:******* * :.*:.:: :: .*::*   *  :*. *****
```

Identity (*): Strongly similar (:) : Weakly similar (.)

Fig.2 Multiple sequence alignment of Fusicoccin by MUSCLE

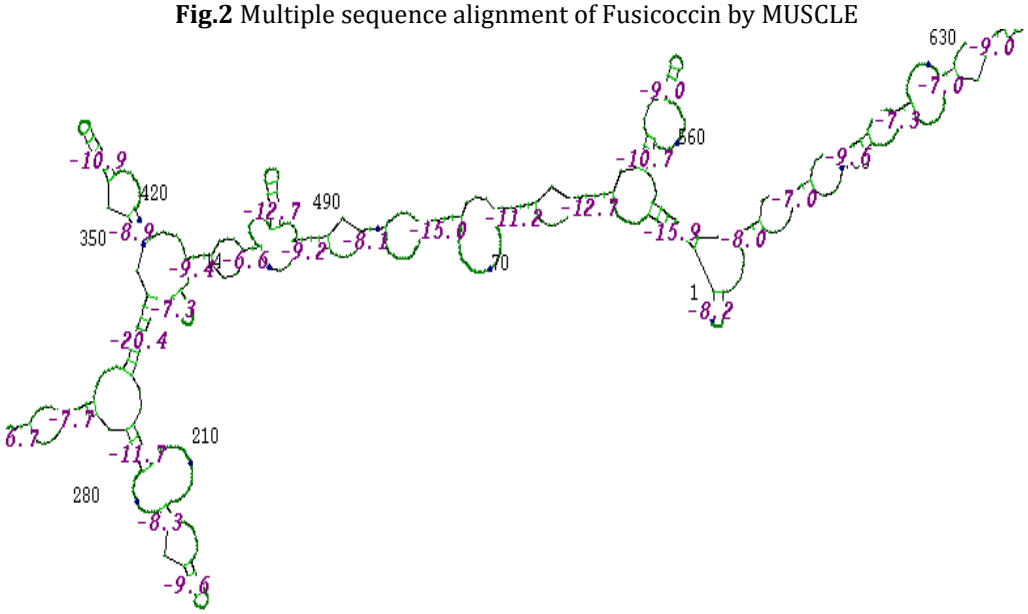
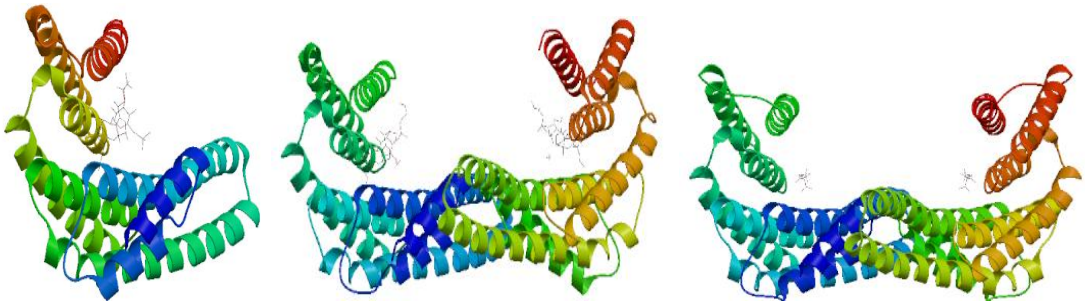


Fig. 3 RNA structure prediction: (ref CAA 65148.1)

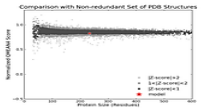


Model 1 Model 2 Model 3

Fig.4: Swiss Models: (ref CAA 65148.1)

Model #01	File	Built with	Oligo-State	Ligands	GMQE	QMEAN4
	PDB	ProMod Version 3.70.	MONOMER	1 x FSC: FUSICOCCIN;	0.90	-0.75

QMEAN4	-0.75
Cβ	0.48
All Atom	2.00
Solvation	1.50
Torsion	-1.83



Template	Seq Identity	Oligo-state	Found by	Method	Resolution	Seq Similarity	Range	Coverage	Description
2o98.1.B	88.51	hetero-oligomer	BLAST	X-ray	2.70Å	0.56		0.92	14-3-3-like protein C

Ligand	Added to Model	Description
FSC	✓	FUSICOCCIN
FSC	✕ - Clashing with protein.	FUSICOCCIN
SO4	✕ - Not biologically relevant.	SULFATE ION
SO4	✕ - Not biologically relevant.	SULFATE ION

```
Target    MASPREENVYMANVADEAERYEEMVEFMEKVVAALNGEELTVEERNLLSVAYKNVIGARRASWRIISSIEQKEESRGNED
2o98.1.B  ----REENVYMAKLAEQAERYEEMVEFMEKVSNSLGSEELTVEERNLLSVAYKNVIGARRASWRIISSIEQKEESRGNEE

Target    HVASIKKYRSQIENELTSICNGILKLLDSKLIGSAATGDSKVFYLKMKGDYYRYLAEFKTGTERKEAAENTLSAYKSAQD
2o98.1.B  HVNSIREYRSKIENELSKICDGILKLLDAKLIPSAASGDSKVFYLKMKGDYHRYLAEFKTGAERKEAAESTLTAYKAAQD

Target    IANGELAPTHPIRLGLALNFSVFYYEILNSPDRACNLAKQAFDEAIAELDTLGEESYKDSTLIMQLLRDNLTLWTSDMQD
2o98.1.B  IATTELAPTHPIRLGLALNFSVFYYEILNSPDRACNLAKQAFDEAIAELDTLGEESYKDSTLIMQLLRDNLTLWTSDMQ-

Target    DGTDEIKEPSKADNE
2o98.1.B  ---------------
```

Fig.5 Model building using SWISS model (ref CAA 65148.1)

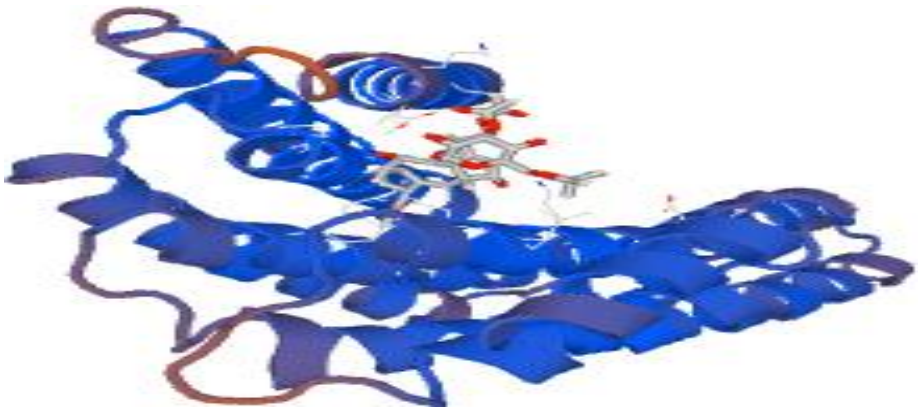


Fig.63D view of the structure with ligand binding (ref CAA 65148.1)

Sequence Homology Analysis: The sequence homology is analyzed by MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/)(Fig:2).

RNA structure prediction: The protein sequences were reverse transcribed to DNA using Sequence manipulation suite (SMS) (http://www.bioinformatics.org/sms2/rev_trans.html). The reverse transcribed DNA was converted to RNA using transcriptional and translational tool (http://www.attotron.com/cybertory/analysis/trans.html). RNA structure was predicted using (http://www.genebee.msu.su/services/rna2_reduced.html) (Table-8 & Fig:3).

Swiss model: (http://swissmodel.expasy.org/) SWISS-MODEL is a fully automated protein structure homology-modelling server, accessible via the ExPASy

web server, or from the program DeepView (Swiss Pdb-Viewer). The purpose of this server is to make Protein Modelling accessible to all biochemists and molecular biologists worldwide. (Fig: 4, 5&6).

Results and discussion

Amino acid composition

The results of Primary sequence analysis of 10 fusicoccin proteins analyzed by CLC work bench revealed the sequence length ranging from 229 -261 (Table 1). The abundant amino acids were glutamic acid, leucine, alanine, and lysine which are tabulated in CLC work bench (Table 2). The most abundant amino acid being Leucine and Glutamic acid with 0.115% in NP_001234007.1 and NP_001234278.1 respectively, and least being tryptophan with 0.008% in majority of the sequences under consideration.

Primary sequence analysis

The result of primary structure analysis suggests that most of the fusicoccin are hydrophobic in nature due to presence of high non-polar residues content (Table-3).The hydrophobic residues are alanine, phenylalanine, glycine, isoleucine, leucine, methionine, proline, valine, tryptophan. The hydrophilic residues are cysteine, asparagines, glutamine, serine, threonine and tyrosine. The highest hydrophobic count is seen in P93206.2|14331_SOLLC.

Physico-chemical parameters:

The average molecular weight of fusicoccin was found in between 25791.8-29431.9 Da. Extinction co-efficient of fusicoccin at 280nm ranged from 24535 –30495M^{-1} Cm^{-1}

Isoelectric point is the pH at which the surface of protein is covered with charge but net charge of protein is zero. pI of fusicoccin was found to be acidic in nature. Computed isoelectric point of proteins > 7 soluble in basic buffers. Isoelectric point is predicted ranges from 4.61-4.96 (Table 4) .Useful for developing buffer system for purification of proteins.

The Aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chains which ranges from 79.8 to 92.58. Which infers positive factor for thermostability (ZIkai , 1980).

The Grand Average hydropathy (GRAVY values) showed that all proteins are hydrophilic ranging from-0.269 to -0.659 supports the soluble nature of fusicoccin proteins. Though, it can play a role in substrate recognition. Here the protein sequences showing negative that indicates stability of the protein. In particular, hydrophobic amino acids can be involved in binding/recognition of ligands.

A protein whose instability index is smaller than 40 are predicted as stable, and a value above 40 predicts that the protein may be unstable, only one fusicoccin sequence(p93206.2I14331_SOLLC) showed stability with instability index of 37.97 (Guruprasad *et al*, 1990)(Table 4)

Secondary structure prediction

SOPMA was employed for calculating the secondary structural features of the selected protein sequences considered in this study. The predicted secondary structural information of the protein was considered to improve the target-template alignment and for building conformations for 3D model of the fusicoccin. This method calculates the content of α-helix, β-sheets, turns, random coils and extended strands. SOPMA is a neural network based methods; global sequence prediction may be done by this sequence method (Mugilan, 2010).

The secondary structure of fusicoccin contain more of helices and coils, where helix range from 54.79%-68.67% and coils from 14.86%-25.67%. Sheets and turns are in less abundance with maximum percentage being 14.56 and 4.98 respectively.

Being hydrophobic, Leucine prefers to be buried in protein hydrophobic cores. Proline has a special property of creating links in polypeptide chains and disrupting ordered secondary structure.

The consequence in which most of the amino acid side chains of trans membrane segments is non-polar (e.g. Ala, Val, Leu, Ile, Phe) and the very polar CO-NH groups (peptide bonds) of the polypeptide backbone of trans membrane segments which participates in hydrogen bonding (H-bonds) in order to lower the cost of transferring them into the hydrocarbon interior. This H-bonding is most easily accomplished with alpha-helices for which all peptide bonds are H-bonded internally. On this basis we can say this may act as a neurotoxic drug target (Table-5)

SMART Analysis

Many proteins are multidomain in character and possess multiple functions that often are performed by one or more component domains. A Web-based tool (SMART) has been designed that makes use of mainly public domain information to allow easy and rapid annotation of signaling multidomain proteins. The tool contains several unique aspects, including automatic seed alignment generation, automatic detection of repeated motifs or domains, and a protocol for combining domain predictions from homologous subfamilies. The ability of SMART to annotate single sequences or large datasets is exemplified by the cases described in fusicoccin. Expect value (E) a parameter that counts the number of hits one can "expect" to see by chance for a database of a particular size. It decreases exponentially as the Score (S) of the match increases. Here it is in the expected range 1.24 to 5.63 (Table 6).

Support vector machines (SVM) method

Support vector machines method for the classification of proteins with diverse sequence distribution.

SVMProt shows a certain degree of capability for the classification of distantly related proteins and homologous proteins of different function and thus may be used as a protein function prediction tool that complements sequence alignment methods. It has been employed in protein studies including protein–protein interaction prediction, fold recognition, solvent accessibility and structure prediction . The prediction accuracy ranges from 1.1 to 99% in this study. Thus SVM classification of protein functional family,a potentially developed into a protein function prediction tool to complement methods based on sequence similarity and clustering.

Based on the Classification of proteins of our interest and its values, we predict that, these proteins contains all DNA binding site, Metal binding site, Zinc binding site, Mg binding site and MTB binding site .

Transmembrane prediction

TMbase is a database of transmembrane proteins and their helical membrane- spanning domains. Possible transmembrane helices, of the accession number ref CAA 65148.1, the sequence positions inside to outside 1 helices is found and outside to inside 1 helices is found. Transmembrane topology suggestions are purely speculative and should be used with extreme caution since they are based on the assumption that all transmembrane helices have been found. In most cases, the prediction plot (Fig 1) that is created should be used for the topology assignment of unknown proteins.

inside->outside | outside->inside

(172- 188 (17) 114) |(169- 188 (20) 366 ++)

Helices shown in brackets are considered insignificant.

Sequence homology Analysis

Multiple Sequence alignment by MUSCLE online tool. The identification of catalytic residues is a key to understanding the function of enzymes. With the information from other functionally similar sequences with known crystallographic structures we can identify the key catalytic residues. Homology sequences revealed significant conserved and semi conserved regions represented as Identity (*): Strongly similar: (:), Weakly similar (.) as shown in (Fig 2).

RNA structure prediction

RNA is now appreciated to serve numerous cellular roles, and understanding RNA structure is important for understanding a mechanism of action. This contribution discusses the methods available for predicting RNA structure. Secondary structure is the set of the canonical base pairs and secondary structure can be accurately determined by comparative sequence analysis. Secondary structure can also be predicted. The most commonly used method is free energy minimization. The accuracy of structure prediction is improved either by using experimental mapping data or by predicting a structure conserved in a set of homologous sequences. Additionally, tertiary structure, the three-dimensional arrangement of atoms, can be modeled with guidance from comparative analysis and experimental techniques. New approaches are also available for predicting tertiary structure.

SWISS Model

The 3D structure analysis of Fusicoccin were done by using SWISS-MODEL automated modeling server, the three models are shown (Fig 4). Template selection, alignment and model building are done completely automated by the server of the ID number CAA 65148.1**.** Predicting the protein 3D structures by this method are used which implements the four steps of the homology modeling approach (Fig 4).

a. Template searching to identify the structure homology: Template search with Blast and HHBlits has been performed against the SWISS-MODEL template library (SMTL). The target sequence was searched with BLAST (Altschul *et al*., 1997) against the primary amino acid sequence contained in the SMTL. A total of 93 templates were found. An initial HHblits profile has been built using the procedure outlined in (Remmert, *et al*., 2011), followed by 1 iteration of HHblits against NR20. The obtained profile has then be searched against all profiles of the SMTL. A total of templates were found.

*b. Template selection***:** For each identified template, the template's quality has been predicted from features of the target-template alignment. The templates with the highest quality have then been selected for model building.

c. Model building: Models are built based on the target-template alignment using Promod-II. Coordinates which are conserved between the target and the template are copied from the template to the model. Insertions and deletions are remodeled using a fragment library. Side chains are then rebuilt. Finally, the geometry of the resulting model is regularized by using a force field. In case loop modelling with ProMod-II (Guex and Peitsch, 1997) does not give satisfactory results, an alternative model is built with MODELLER (Sali and Blundell,1993) (Fig 5).

*d. Model quality estimation***:** The global and per-residue model quality has been assessed using the QMEAN scoring function (Benkert *et al*, 2011)

Ligand Modeling

Ligands present in the template structure are transferred by homology to the 3D model of the ID

number (ref CAA 65148.1)when the following criteria are:

(a) The ligands are annotated as biologically relevant in the template library,
(b) The ligand is in contact with the model,
(c) The ligand is not clashing with the protein,
(d) The residues in contact with the ligand are conserved between the target and the template.If any of these four criteria is not satisfied, a certain ligand will not be included in the model. The model (Fig 6) summary includes information on why and which ligand has not been included.

Conclusion

Fusicoccin is the widely studied phytotoxin, which causes endogenous modulation of proton pump activity in green plants. In the present study, the sequence and structure analysis of Fusicoccin protein was done by various tools and software's. Based on the findings it could be concluded that further characterization of FC is novel and will be important for evaluating relatedness of the protein with the eukaryotic protein called 14-3-3. The wide range of action of fc protein can be correlated to the presences of fusicoccin binding protein (FBP) present in most of the plant species studied. FBP is member of 14-3-3 superfamily.

The present analysis employs members of NP_ 001296299.1 , NP_ 001234267.1 , NP_ 001234097.1 , NP_ 001234007.1 , NP_ 001234677.1 , NP_ 001234278.1 , NP_ 001234637.1 , NP_ 001234272.1 , CAA 65148.1 and P93206.2|14331_SOLLC , selected from NCBI database showing high conservation which suggests their functional similarity. Physicochemical properties of fusicoccin were analyzed by using ProtParam. Physicochemical characterization studies give more insight about the properties such as Molecular Wt, pI, AI, GRAVY and Instability Index that are essential and vital in providing data about the proteins and their properties

From the above mentioned results, it is clear that the 3D structure of the fusicoccin protein is similar to the 14-3-3 proteins, a class of mammalian brain proteins initially described as regulators of neurotransmitter synthesis and protein kinase C inhibitors and shows large number of conserved regions. The physico chemical properties of the fc protein have shown the wide use of protein in the tomato plant in the root development and the novel properties of the protein is that it can be used for the almost all the green plants.

It is becoming clear that fusicoccin may have many important functions, any or all of which might contribute to its stomatal opening. Although significant progress has been made towards elucidating its role in stomatal opening, further work will be required in order to fully understand how fusicoccin is regulated and role in increase in fresh mass of roots. Further functions of FC protein include role in seed germination, seed dormancy.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997), Gapped BLAST and PSI-BLAST, a new generation of protein database search programs, *Nucleic Acids Res*, 25, pp 3389-3402.

Ashokan, K.V., Mundaganur, D.S, and Mundaganur, Y.D. (2011), Catalase: Phylogenetic characterization to explore protein cluster. *Journal of research in Bioinformatics.* 1, pp 001-008.

Benkert, P., Biasini, M. and Schwede, T. (2011), Towards the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, 27,pp 343-350.

Brown, P. H., Outlow W H, (1982), Effect of Fusicoccin on Dark '4CO2 Fixation by Viciafaba Guard Cell Protoplasts. *Plant Physiol*. 70, pp 1700-1703.

Dohrman U**,** Hertel R, Pesci P, Cocucci S**,** Marrè E**,** Randazzo G**,** Ballio A (1977), Localization of in vitro binding of the fungal toxin fusicoccin to plasma-membrane rich fractions from maize coleoptiles. *Plant Sci Lett*9**,** pp291-299

Eisenhaber, F., Imperiale, F., Argos, P. and Froemmel, C. (1996), Prediction of secondary structural content of proteins from their amino acid composition alone. I. New analytic vector decomposition methods. *Proteins: Struct. Funct. Design*.25, pp 157-168.

Gill, S.C. and Von Hippel, P.H. (1989), Calculation of protein extinction coefficients from amino acid sequence data. *Anal.Biochem*.182, pp 319-326

Guex, N. and Peitsch, M.C. (1997), SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis,* 18, pp 2714-2723.

Guruprasad, K., Reddy, B.V.P. and Pandit, M.W. (1990), Correlation between stability of a 273 protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Prot.Eng*. 4 (2), pp 155-164.

Hofmann, K. and Stoffel, W. (1993), TMbase - A database of membrane spanning proteins segments, *Biol. Chem.Hoppe-Seyler* 374, pp 166

Jahn T, Fuglsang AT, Olsson A, Bru Èntrup IM, Collinge DB,Volkmann D, Sommarin M, Palmgren MG, Larsson C. 1997,The 14-3-3 protein interacts directly with the C-terminal region of the plant plasma membrane H+-ATPase. The Plant Cell 9, pp 1805 ± 1814.

Johansson, F., Sommarin, M., and Larsson, C. (1993), Fusicoccin activates the plasma membrane H+-ATPase by a mech- anism involving the C-terminal inhibitory domain, *Plant Cell* 5, pp 321-327

Letunic, I., Doerks, T. and Bork, P. (2009) SMART 6: recent updates and new developments. *Nucleic Acids Res.* 37, pp 229/32

Madhu, S. and Mahesh, P. (2015), Sequence analysis of Semaphorin in tumor progression: An Insilico approach, *International Journal of Asian Academic Research of Multidisciplinary* ,1(29), pp 407-423.

Mahesh, P., Divya, P., Akshatha, M., Prathima, R. and Lava Kumar, C. (2015), Insilico characterization and phylogenetic analysis of novel probiotic bacteria in honey bees. *International Journal of Asian Academic Research of Multidisciplinary* ,1(32), pp 337-357.

Malerba M, Crosti P, Armocida D, Bianchetti R (1995), Activation of ethylene production in Acer pseudoplatanus L. cultured cells by fusicoccin, J Plant Physiol 145, pp 93±100

Marre`, E. (1979) Fusicoccin: a tool in plant physiology, Annu. Rev, *Plant Physiol*. 30, pp 273–288.

Marre E (1980) Fusicoccin: mechanism of action on electrogenic H+ extrusion. In RM Spanswick, WJ Lucas, J Dainty, eds, Plant Membrane Transport: Current Conceptual Issues, *Elsevier/North-Holland Biomedical Press*, New York, pp 227-240

Moyen c, Johannes E (1996), systemin transiently depolarizes the tomato mesophyll cell membrane and antagonizes fusicoccin-induced extracellular acidification of mesophyll tissue*, Plant Cell Environ* 19, pp 464-470

Mugilan, A., Ajitha, M.C., Devi and Thinagar.. (2010) —In silico Secondary Structure Prediction Method (Kalasalingam University Structure Prediction Method) using Comparative Analysis, *Trends in Bioinformatics*. 3(1),pp 11-19.

Oecking, C., Piotrowski, M., Hagemeier, J., and Hagemann, K. (1997), Topology and target interaction of the fusicoccin-binding 14-3-3 homologs of Commelina communis, Plant J. 12, pp 441–453

Praveen Kumar, K.S. and Mahesh pattabhiramaiah. (2015) Sequence analysis of basic phospholipase a2 (neurotoxin) as a potential drug target: an in silico approach. *International Journal of Engineering Research and General Science* Volume 3, Issue 3. pp 1057-1067

Remmert, M., Biegert, A., Hauser, A. and Soding, J. (2012) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment, *Nat Methods*, 9, pp 173-175.

Sali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints, *J Mol Biol*, 234, pp 779-815.

ZIkai, A. (1980), Thermo stability and aliphatic index of globular proteins, *J. Biochem*. 88,pp 1895-1898.