American Journal of

Biochemistry and Molecular Biology

ISSN 2150-4210

Academic
Journals Inc.

www.academicjournals.com

American Journal of Biochemistry and Molecular Biology 3 (2): 188-201, 2013
ISSN 2150-4210 / DOI: 10.3923/ajbmb.2013.188.201

Homology Modeling of a Fruit Ripening Specific Plant MADS–box Factor

[1]Sudip Kumar Sinha and [2]Dibyendu N. Sengupta
[1]Department of Botany, Victoria Institution (College), 78B, A.P.C. Road, Kolkata-700009, India
[2]Division of Plant Biology, Bose Institute, 93/1, A.P.C. Road, Kolkata-700009, India

Corresponding Author: Sudip Kumar Sinha, Department of Botany, Victoria Institution (College), 78B, A.P.C. Road, Kolkata-700009, India

ABSTRACT

A MADS-box (Minichromosome maintenance-1, Agamous, Deficiens and Serum response factor) transcription factor namely SlMADS RIN (*Solanum lycopersicum* MADS Ripening inhibitor) seems to act as global regulator in climacteric fruit ripening process of tomato. Structure modeling of any plant MIKC (MADS-box, I-box, K-box and C-box) -type MADS-box factor were unknown till date, and the present study is an approach towards this direction. The template search of SlMADS RIN was performed by PSI BLAST (Position-Specific Iterative Basic Local Alignment Search Tool). Initial model was built with the help of MODELLER 9v4 package. The predicted 3D structure for SlMADS RIN protein was further validated by Ramachandran plot analysis using the PROCHECK tool. The submitted sequence of SlMADS RIN protein to PSI BLAST tool identified only one region (1-73 amino acids). DOPE (Discrete Optimized Protein Energy) score analysis revealed that the modeled structure showed overall lower DOPE score value (-3968.569336). Ramachandran plot analysis revealed that 94.1% residues were in favored region and 05.9% residues were in allowed region. Result revealed that the SlMADS RIN protein structure under study deviate largely by sequence with the known MADS-box, except the N-terminal 74 amino acid. Further, the side chain and loops of SlMADS RIN showed <1Å (0.233 Å) root mean square deviation. Thus, it can be concluded that this is the first report on prediction of three dimensional models for SlMADS RIN and this modeled structure can be used to predict the molecular function of the protein.

Key words: Homology modeling, MADS-box, SlMADS RIN, tomato

INTRODUCTION

The post harvest physiological, biochemical and molecular changes in edible fruits are mainly responsible for ripening. Either the developmental program alone or along with additional events induced by ethylene, the only gaseous plant hormone, triggers the ripening process of fruits (Giovannoni, 2004). Tomato (*Solanum lycopersicum* L.) is a model plant for analysis of ripening in fruits. It is because of its diverse germplasm, high density physical map and large number of EST (Expressed Sequence Tag) collection. Moreover, the availability of molecular tools, genome sequencing project and efficient transformation procedure in tomato are also mention worthy (Cara and Giovannoni, 2008). A number of tomato ripening mutants have been characterized till date, among which rin (ripening inhibitor) mutant recently characterized extensively. Mapping and positional cloning of rin mutant revealed that MADS-box (Minichromosome maintenance-1, Agamous, Deficiens and Serum response factor) genes act as regulators of fruit ripening and a

specific MADS-box protein seems to be involved in ripening process (Vrebalov *et al*., 2002). Further the protein SlMADS RIN (*Solanum lycopersicum* MADS Ripening inhibitor) was expressed specifically in ripe tomato fruits and its expression is not significantly influenced by ethylene. But the SlMADS RIN is required to initiate climacteric respiration and associated ethylene biosynthesis in addition to ripening factors that cannot be complemented by external ethylene application. Consequently, SlMADS RIN is upstream of ethylene in the regulatory cascade. *In vivo* assays revealed that RIN binds to the *cis*-acting element of *SlACS2* (*Solanum lycopersicum* ACC Synthase2) and may represent a global developmental regulator of ripening, potentially shared among climacteric and non-climacteric species (Ito *et al*., 2008).

Crystal and NMR (Nuclear Magnetic Resonance) structure study of several animal and fungal MADS-box transcription factors have already been resolved, but no such study for any plant MIKC-type (MADS-box, I-box, K-box and C-box) MADS-box factor known till date. The approach of homology modeling has been used in the present study to resolve the structure of MADS-box domain of SlMADS RIN protein.

MATERIALS AND METHODS

Domain identification: Domain identification of SlMADS RIN protein was carried out by submitting the SlMADS RIN protein sequence to Conserved Domain Database (CDD) of NCBI (http://www.ncbi.nlm.nih.gov/cdd). Protein-protein queries submitted to NCBI's BLAST (Basic Local Alignment Search Tool) search service are scanned for the presence of conserved domains by default.

Template search and sequence alignment: The complete SlMADS RIN protein sequence, comprising of 242 amino acid residues, was subjected to PSI BLAST (Position-Specific Iterative BLAST) (Altschul *et al*., 1997) in order to find homologous sequences having close similarity with the domain regions suggested by CDD. Protein Data Bank (PDB) was chosen for homology searching using BLOSUM 62 matrix with gap costs of 11 and 1. The threshold E-value was set at 0.001 and all other parameters of PSI BLAST tool were set at default values. The PDB codes of all those template sequences that got aligned at different local regions of the query SlMADS RIN protein and with a better E-value than the set threshold were noted along with the % identity and % gap values. In this program, a profile is constructed from a multiple alignment of the highest scoring hits in an initial blast search. The position specific scoring matrix was generated by calculating position-specific scores for each of the positions in the alignment. The profile was used to perform a second, third, etc., BLAST search, and the results of each iteration was used to refine the profile. This iterative searching strategy resulted in increased sensitivity.

The common domains of SlMADS RIN protein identified by CDD of NCBI were submitted to FUGUE program (Shi *et al*., 2001). FUGUE quantitatively assesses sequence similarity in terms of 3-dimensional (3D) structures. It defines a structural environment in terms of main-chain conformation, secondary structure, solvent accessibility, and also H-bonding status.

In order to detect the conserved amino acids residues in domain of SlMADS RIN protein identified by CDD of NCBI, fragment of SlMADS RIN protein and the corresponding PDB templates identified by PSI BLAST analysis were submitted to CLUSTALW tool (Thompson *et al*., 1997). All the parameters of CLUSTALW were set at default values.

Tertiary structure prediction: The tertiary structure for the domain of SlMADS RIN protein, identified by CDD of NCBI were predicted by molecular modeling software, MODELLER 9v4 (Sali and Blundell, 1993). MODELLER is a computer program that models protein structure by satisfaction of spatial restraints.

Alignments between the SlMADS RIN protein (1-74 amino acids) and the corresponding four templates were carried out using 'align2d_mult' command. Final modeling was performed using 'model_mult' command to model SlMADS RIN protein 1-74 amino acid. Among the five modeled structures that were generated for SlMADS RIN protein (1-74 amino acid) by the MODELLER program, the structure with the lowest probability density function or MODELLER objective function was considered for subsequent analysis.

Several loops in SlMADS RIN protein 1-74 amino acid were identified, by submitting the sequence information to GOR4 server (Garnier *et al*., 1996). MODELLER (9v4) package (Fiser *et al*., 2000) was used to build the loop that was identified by GOR4. Loop modeling of C_{α}-backbone for all the loops of SlMADS RIN protein was done using 'loop_refine' of MODELLER software. An initial loop conformation is then generated by simply positioning the atoms of the loop with uniform spacing on the line that connects the main-chain carbonyl oxygen and amide nitrogen atoms of the N- and C-terminal anchor regions, respectively. Number of such loop models are generated, each taking the initial loop conformation and randomizing it by + or -0.5 Å in each of the Cartesian directions. The model is then optimized that relies on an atomistic distance-dependent statistical potential of mean force for non-bond interaction (Melo and Feytmans, 1997). Structural superposition of backbone before and after loop modeling was done using 'MatchMaker' command of UCSF Chimera (Pettersen *et al*., 2004).

Side chain residues of loop refined backbone model were modeled by submitting to SCWRL3 software (Canutescu *et al*., 2003) with all atom side chain modeling command.

Energy minimization and force field application of the modeled structure of SlMADS RIN protein was done with the help of SWISS-pdb Viewer (Guex and Peitsch, 1997), using both Steepest Descent and Conjugate Gradients of AMMBER97 force field up to 500 iteration each.

Evaluation and calculation of root mean square deviation: Once a final model was selected, '*assess_dope*' command of MODELLER was used to evaluate the model fold. To calculate root mean square deviation of the modeled structure from the template 'MatchMaker', command of UCSF Chimera was used. When comparing protein models with DOPE (Discrete Optimized Protein Energy), the model with the lowest DOPE score was considered the most favorable protein model.

Verification of tertiary structure: The predicted 3D structure for SlMADS RIN protein 1-74 amino acid was further validated by Ramachandran plot analysis using the PROCHECK tool (Laskowski *et al*., 1993). The overall G-value for the predicted structure was obtained by submitting the best predicted MODELLER output file to PROCHECK (Laskowski *et al*., 1993).

Classification: CATH (Class, Architecture, Topology and Homologous super family) server available at http://www.cathdb.info/index.html (Orengo *et al*., 1997) was utilized for classifying the SlMADS RIN protein into Class, Fold etc.

Active site identification: Accessible surface area of modeled SlMADS RIN protein (1-74 amino acid) was determined using 'calculate accessible surface (in detail) and 'solid surface (build)'

commands of NOC 3.01 (http://noch.sourceforge.net/). Active pockets of modeled SlMADS RIN protein was determined using 'Pocket Prediction' command of Pocket Picker plug-in (Weisel *et al.*, 2007) of PyMOL visualization software (http://pymol.sourceforge.net/).

RESULTS AND DISCUSSION

Template searching: The CDD of NCBI revealed only two conserved domains for SlMADS RIN protein, namely MADS_MEF2 like and K-box super family (Fig. 1). The submitted sequence of SlMADS RIN protein to PSI BLAST tool, for retrieving the homologous sequences deposited in PBD database, identified only one region (1-73 amino acid) (Fig. 2). As no new template sequences were identified after the third iteration, the number of iterations in the PSI BLAST run was restricted to three. The alignment data showed that only 1-74 region of target sequence was aligned with the four template sequences like (1TQE, 1EGW, 1N6J and 1C7U), and showed higher coverage of target sequence and lower E-value (Table 1). Template 1TQE showed 54% identity with the target sequence having the E-value 1e-29. Template 1EGW showed 57% identity with the target sequence having the E-value 1e-29. Template 1N6J showed 53% identity with the target sequence having the E-value 4e-29, whereas, Template 1C7U showed 57% identity with the target sequence having the E-value 1e-27. Again, all four template sequences showed only 1% gap (Table 1). Moreover, all four templates were annotated as Myocyte Enhancer Factor-2 or -2a (Table 1). Template 1K6O and

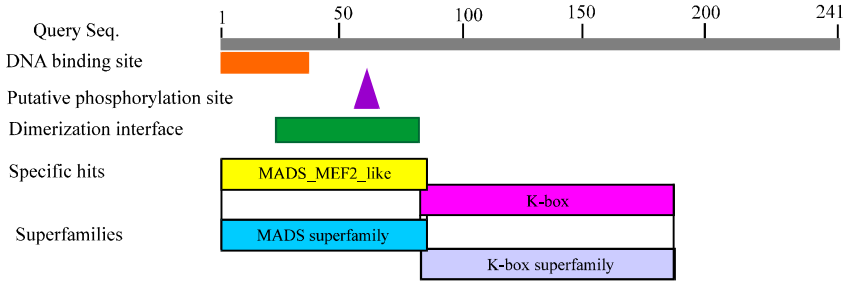


Fig. 1: Schematic diagram of different domain regions of SlMADS RIN (Solanum lycopersicum MADS Ripening inhibitor) protein identified by Conserved Domain Database of NCBI using full length protein sequence of SlMADS RIN as query, MADS: Minichromosome maintenance-1, Agamous, Deficiens and Serum response factor; MEF2: Myocyte Enhancer Factor-2

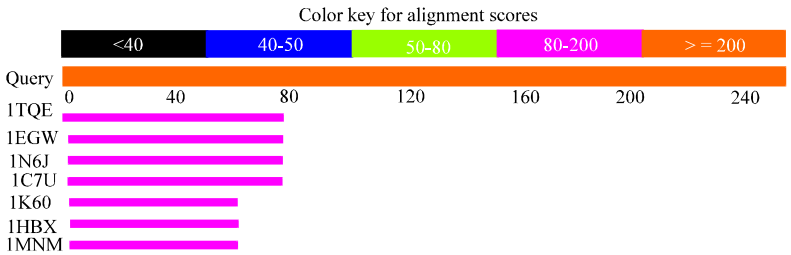


Fig. 2: Schematic diagram to show different PDB (Protein Data Bank) hits in respect to SlMADS RIN (*Solanum lycopersicum MADS Ripening inhibitor*) amino acid sequence as retrieved after seventh iteration of PSI (Position Specific Iterative)-BLAST search

Table 1: Summary of the best template sequence profile that was generated at the end of seventh iteration of PSI BLAST* analysis using SlMADS RIN*

SlMADS RIN region	PDB* ID of Template	Region of template sequence aligned	Identity (%)	Gap (%)	e-value	Annotation of template
1-74	1TQE chain P, Q, R and S	1-74	54	1	1e-29	Myocyte Enhancer Factor-2
	1EGW chain A, B, C, and D	2-74	57	1	1e-29	Myocyte Enhancer Factor-2[a]
	1N6J chain A & B	2-74	53	1	4e-29	Myocyte Enhancer Factor-2
	1C7U chain A and B	2-74	57	1	1e-27	Myocyte Enhancer Factor-2[a]
	1K6O chain B and C	2-59	43	0	3e-21	Serum responsive factor
	1MNM chain A and B	3-59	47	0	1e-20	MCM1

*PSI BLAST: Position Specific Iterative Basic Local Alignment Search Tool, PDB: Protein Data Bank, SlMADS RIN: Solanum lycopersicum MADS Ripening inhibitor

1MNM showed lower degree of target sequence coverage (2-59 amino acid residue and 3-59 amino acid residue, respectively) and higher E-value (3e-21 and 1e-20, respectively) and thus discarded from further analysis.

Generally homology modeling is performed with the protein targets which share >30% amino acid sequence identity (Baker and Sali, 2001). It is because of that the reliability of the sequence alignment between target and template decreases rapidly within 30% sequence identity. Although, >28% template-target identity can produce significantly reliable model in some cases (Guleria and Yadav, 2013; Smith and Plazas, 2011). Medium accurate models, resolved with 30-50% of sequence similarity, have nearly 85% of their Cα atoms within 3.5 Å. These models can be used in variety of applications, like designing site directed mutants, screening of combinatorial small molecule database, etc. Top accurate models, fetched on sequence identities >50%, usually have structures comparable to 3 Å resolution X-ray structures. It can be used for more reliable calculations as ligand docking, drug design, etc. However, sequence identities >90% can be used to describe the active site (Marsden and Orengo, 2008). Profile-profile method was used in this study because it performed at least 30% better than standard sequence-profile methods (Ohlson *et al*., 2004).

In PDB, 1C7U, 1EGW 1N6J, 1MNM, 1K6O and 1SRS MADS transcription factor entries are present. Most *Arabidopsis* MADS-box transcription factors only have obvious similarity with 1MNM, 1EGW (Yeast MEF2A) (Santelli and Richmond, 2000). The MEF2A core protein (amino acid residues 2-78) is folded in the same three-layer organization observed for SRF and MCM1 (Mo *et al*., 2001; Tan *et al*., 2000). The output of FUGUEE program for SlMADS RIN showed similar SRF template as suggested by PSI-BLAST tool (Table 2). The Z-score of the templates was 19.40 which have more than the cut-off value 6. It suggests a high certainty for this template (Table 2). Although, the other templates suggested by FUGUEE showed Z-score value lower than the cut-off value of 6 and hence not taken for this study.

Multiple sequence alignment of N-terminal portion (1-74 amino acid) of SlMADS RIN protein with the templates suggested by PSI-BLAST by CLUSTALX v 1.8 showed that very high conserve-ness. But in a few cases amino acids were substituted by similar type of amino acids (Fig. 3). Based on the length, percent sequence identity, percentage gaps and the E-value of the SlMADS RIN protein that was aligned with various template sequences of PSI-BLAST, 1TQE, 1EGW, 1N6J and 1C7U were selected as the template sequences.

Table 2: Summary of structural comparison by FUGUEE program for SlMADS RIN* protein (1-74 amino acid)

Profile Hit	PLEN*	RAWS*	RVN*	Z score*	ZORI*	AL*	
SRF-TF	85	47	193	19.40	19.87	02	Certain
hs1z56a	77	-68	50	4.64	4.49	02	Likely
hsd1dia1	166	-51	101	4.12	4.87	00	Likely
hs1ho2a	20	-190	17	4.07	3.92	02	Likely
hs2jeea	78	-128	13	3.97	4.74	02	Marginal
hs1by0a	27	46	15	3.78	3.63	22	Marginal
hs1q90l	32	41	16	3.65	3.50	22	Marginal
hs1k6ia	318	-273	109	3.48	4.84	00	Guess
hs2hqya	298	-231	60	3.41	4.69	00	Guess
hs3b77a	185	-68	83	3.39	4.44	00	Guess

*SlMADS RIN: *Solanum lycopersicum* MADS Ripening inhibitor, PLEN: Profile Length, RAWS: Raw alignment scroes, RVN: (Raw score)-(Raw score for NULL model), Z-score normalized by sequence divergence, ZORI: Original Z-score (before normalization), AL: Alignment algorithm used for Z-score/alignment calculation

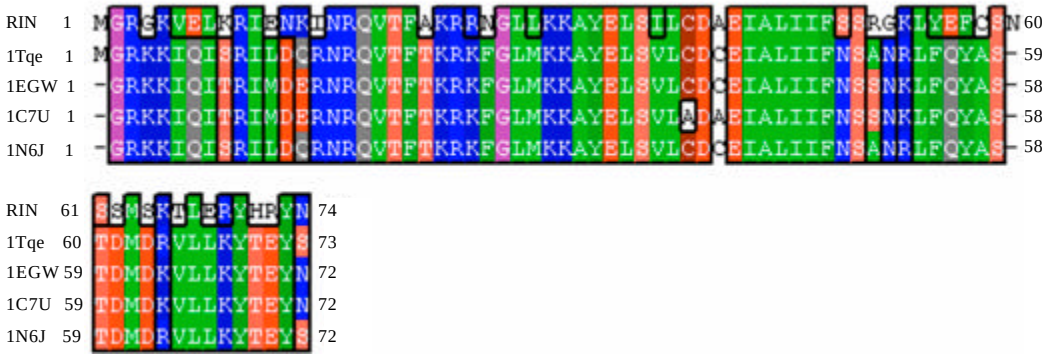


Fig. 3: Multiple sequence alignment of SlMADS RIN (*Solanum lycopersicum* MADS Ripening inhibitor) with template suggested by PSI (Position Specific Iterative)-BLAST using ClustalX program, RIN: Ripening Inhibitor

Model prediction: Alignment between SlMADS RIN (1-74 amino acid) and corresponding four different template sequences shown by 'align2d_mult' of MODELLER was almost similar with the alignment produced by Clustal W (Fig. 3, 4). Among the five 3D models that were generated, the one with lowest modeler objective function was selected as best model (Fig. 5). In the present study MODELLER program was used for predicting the 3D structure of proteins as because the usefulness of this program has been well demonstrated in several studies (Sali and Blundell, 1993).

Model refinement-loop modeling: The loop regions predicted by GOR4 server suggested that five loop regions were present between 1-74 amino acids of SlMADS RIN protein (Fig. 6a). Moreover, from GOR4 server prediction it was found that, SlMADS RIN protein is a helix reach protein. 54.13% of alpha helix was found in the protein whereas, extended strand and random coil comprises of 11.16 and 34.71% respectively (Fig. 6b). The length of the loops in different regions was varied from 1-8. Loop modeling of Cα-backbone for all the loops was done using MODELLER. MODELLER proved to be more accurate for short loops as was observed by Jamroz and Kolinski (2010). Loop modeling further resulted into reduction in DOPE score value

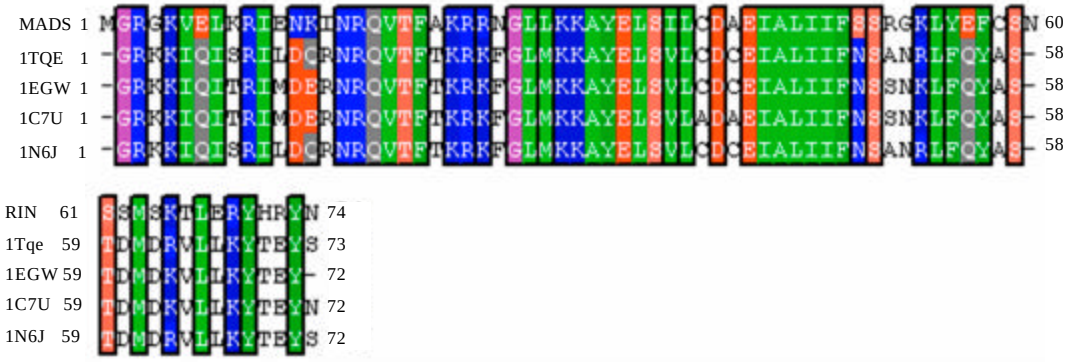


Fig. 4: Multiple sequence alignment of SlMADS RIN (*Solanum lycopersicum* MADS Ripening inhibitor) (1-74 amino acid) with the templates using MODELLER program using 'align2d_mult' command, MADS: Minichromosome maintenance-1, Agamous, Deficiens and Serum response factor

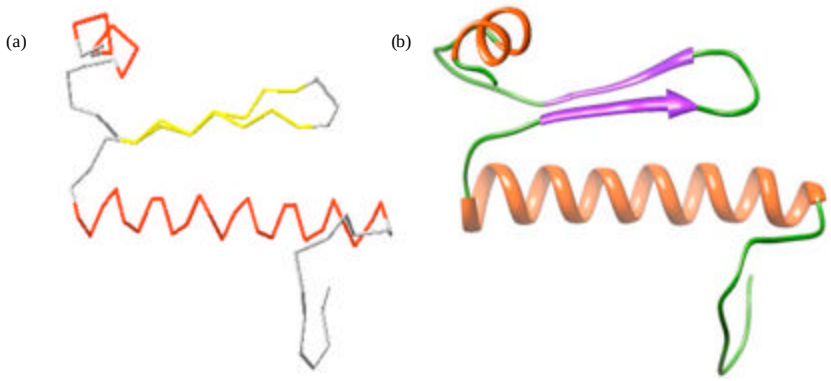


Fig. 5: Initial backbone model of SlMADS RIN (*Solanum lycopersicum* MADS Ripening inhibitor) (1-74 amino acid) as build by '*model_mult*' of MODELLER. (a) Cα-backbone model coloured by secondary structure type, (b) Backbone represented by ribbon and coloured by secondary structure type

(from-5871.276367 to -5704.896973) (Fig. 7). This result supports the fact that in loop modeling, the measure of accuracy is the RMSD (Root Mean Square Deviation) with respect to the main chain atoms after local superimposition of target loop and predicted loop (Joo *et al.*, 2011). Structural superposition of backbone before and after loop modeling, using 'MatchMaker' command of UCSF Chimera, showed that loops of SlMADS RIN were modeled with RMSD of 0.233 Å. The loop region 59-66 was modeled with greater confidence as compared to initial backbone model (Fig. 7). Similarly region 1-5 was also modeled with good confidence (Fig. 7). In this study loops of SlMADS RIN were modeled with RMSD of <1 Å (0.233 Å). This observation was at per with the observation of other investigators (Hooda *et al.*, 2012).

Model refinement-side chain: Side chain of several residues like ILE43, LEU29, LEU54 etc. showed clashes when compared with other residues in the model. That was due to lack of proper rotation of side chain of loop refined model from MODELLER. Thus, the resulted model showed

(a)

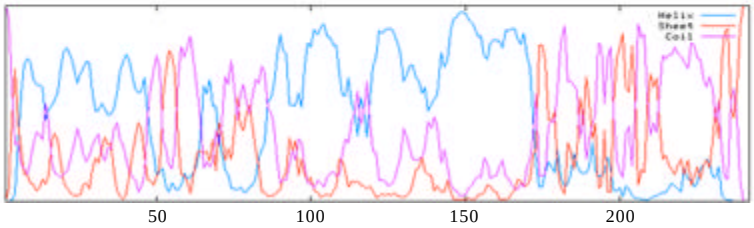
```
MGRGKVELKRIENKINRQVTFAKRRNGLLKKAYELSILCDAEIALIIFSSRGKLYEFCSNSSMSKTLERY
ccccchhhhhhhhhchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccceeeeeccccccccchhhh

HRYNYGTLEGTQTSSDSQNNYQEYLKLKTRVEMLQQSQRHLLGEDLGQLGTKDLEQLERQLDSSLRQIRS
hhccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhhhhhhhhhhhhh

TKTQHILDQLAELQQKEQSLTEMNKSLRIKLEELGVTFQTSWHCGEQSVQYRHEQPSHHEGFFQHVNCNN
hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcceeeecccccccccccceecccccccceeeeeeeeccc

TLPISYGYDNVQPENAAPSTHDATGVVPGUML
ceeecccccccccccccccccccccceeeccceec
```

(b)



Summary of GOR4 prediction:

Alpha helix	(Hh)	: 131	is	54.13%
3_{38} helix	(Gg)	: 0	is	0.00%
Pi helix	(Ii)	: 0	is	0.00%
Beta bridge	(Bb)	: 0	is	0.00%
Extended strand	(Ee)	: 27	is	11.16%
Beta turn	(Tt)	: 0	is	0.00%
Bend region	(Ss)	: 0	is	0.00%
Random coli	(Cc)	: 84	is	34.71%
Am1bigous states (?)		: 0	is	0.00%
Other states		: 0	is	0.00%

Fig. 6(a-b): Secondary structure prediction of SlMADS RIN (*Solanum lycopersicum* MADS Ripening inhibitor) as predicted by GOR (Garnier-Osguthorpe-Robson) server showing position of possible loop regions (c). (a) Residue wise prediction of secondary structural elements present in SlMADS RIN [c = random coil; h = alpha helix; e = extended strand], (b) Graphical and tabular representation of secondary structural elements present in SlMADS RIN

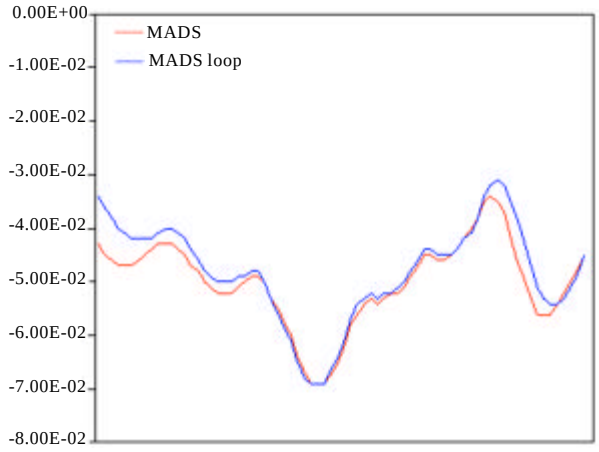


Fig. 7: Comparison of DOPE (Discrete Optimized Protein Energy) value for entire chain of modeled SlMADS RIN (*Solanum lycopersicum* MADS Ripening inhibitor) protein before (MADS) and after loop modeling (MADS loop). X-axis represents number of amino acid residue and Y-axis shows DOPE score, [MADS: Minichromosome maintenance-1, Agamous, Deficiens and Serum response factor]

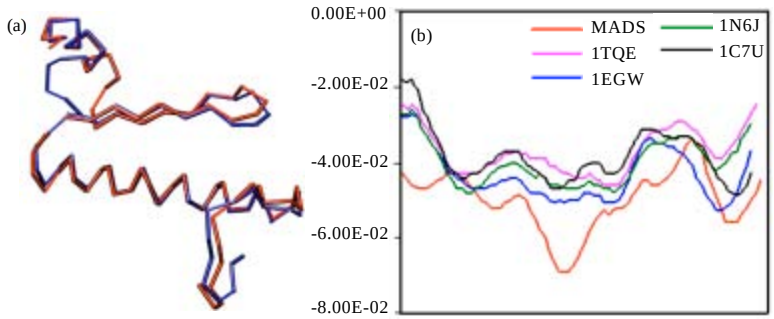
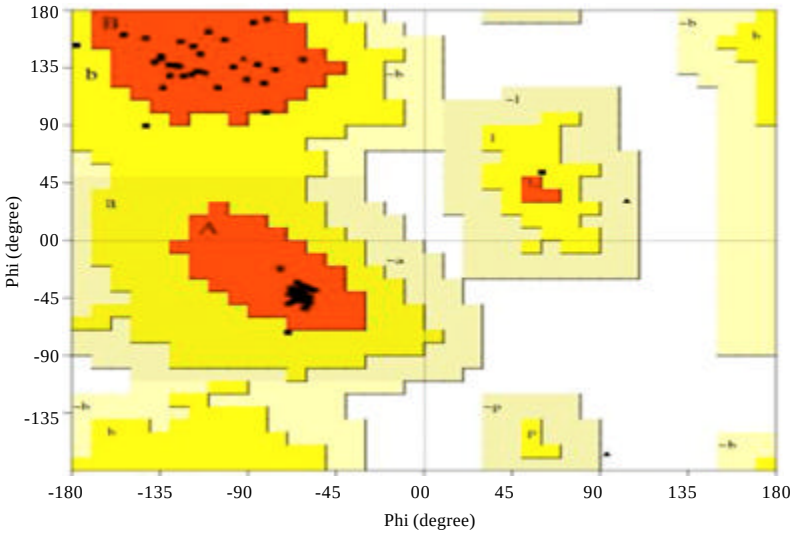


Fig. 8(a-b): Evaluation of the modeled structure of SlMADS RIN (*Solanum lycopersicum* MADS Ripening inhibitor) in respect to target PDB (Potein Data Bank) structures. (A) Cα-backbone superimposition of template 1TQE (red) and target (blue), (b) Comparison of DOPE (Discrete Optimized Protein Energy) score of all four templates used to model with the modeled SlMADS RIN protein. X-axis represents number of amino acid residue and Y-axis shows DOPE score, [MADS: Minichromosome maintenance-1, Agamous, Deficiens and Serum response factor]

proper arrangement of side chains around the space after submitting to SCWRL3 software with all atom side chain modeling command. Although, a clash between LEU29 and PHE57 was still present in the model.

Energy minimization and force field application: After energy minimization, a significant amount of reduction in DOPE score (-3968.569336) was found, which actually indicated the structural stability of the model.

Structure evaluation: DOPE score analysis revealed that the modeled structure, especially the loop region, showed overall lower DOPE score value as compared with all of the templates. This indicates a very good overall conformation of modeled SlMADS RIN protein (1-74 amino acid). Structural superimposition of Cα-backbone of modeled SlMADS RIN protein (1-74 amino acid) with one of the template (1TQE) showed very low RMSD of 1.971 Å (Fig. 8a, b). The RMSD score of 1.971 Å with a DOPE score of -3968.569336 indicated that both template and target proteins have similar folds. That also confirmed the good agreement of the structural model with experimental template as was evident from the study of Hooda *et al.* (2012). Ramachandran plot analysis of modeled SlMADS RIN protein (1-74 amino acid) revealed that 94.1% residues were in favoured region and 05.9% residues were in allowed region (Fig. 9). No residue was found to be Ramachandran outlier. A good quality Ramachandran plot has over 90% in the favoured regions (Prajapat *et al.*, 2011). So, this model proved to be good agreement of the structural model with experimental template. PROCHECK analysis of the submitted modeled structure was done to check the overall stereochemistry. Result revealed the assessment of main chain parameters and showed better overall G-factor and zeta angle standard deviation. G-factor calculated by PROCHECK for the 3D structure of SlMADS RIN protein (1-74 amino acid) was found to be 1.8 which was well above the acceptable threshold of -0.5 (Madhusudhan *et al.*, 2006). But the other parameters showed average result (Table 3). In case of assessment of side chain parameters, all residues showed better resolution in all aspect of stereochemical checks (Table 4). Over all Z-score of PROCHECK G-factor for phi/psi and G-factor for all dihedral angles were 1.06 and 2.66,



Plot statistics		
Residues in most favoured regions (A,B,L)	64	94.1%
Residues in additional allowed regions (a,b,l,p)	4	05.9%
Residues in generously allowed regions (~a,~b,~l,~p)	0	0.0%
Residues in disallowed regions	0	0.0%
	----	--------
Number of non-glycine and non-proline residues	68	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of proline residues	4	
	0	
Total number of residues	----	
	74	

Sterochemical parameter	No. of Deta pts	paramete value	Comparision values Typical	Band	No. of band widths	
a. %-tage residues in A, B, L	48	87.5	83.8	10.0	0.0	Inside
b. Omaga angle st dev	55	5.4	6.0	3.0	-0.2	Inside
c. bad contects/100 residues	0	0.0	4.2	10.0	-0.4	Inside
d. Zeta angle st dev	70	1.1	3.1	1.6	-1.3	Better
e. h-bond energy st dev	37	0.8	0.8	0.2	-0.1	Inside
f. Overall G-factor	114	0.2	-0.4	0.3	1.8	Better

(b)

Sterochemical parameter	No. of Deta pts	Parameter value	Comparision values Typicial Value	band width	No. of band widths from mean	
a. Chi-1 gauche minus st dev	3	6.4	18.1	6.5	-1.8	BETTER
b. Chi-1 trans st dev	22	5.3	19.0	5.3	-2.6	BETTER
c.Chi-1 gauche plus st dev	41	8.6	17.5	4.9	-1.8	BETTER
d. Chi-1 poiled st dev	66	8.5	18.2	4.8	-2.0	BETTER
e. Chi-2 trans st dev	33	9.0	20.4	5.0	-2.3	BETTER

(c)

Fig. 9: Ramachandran plot analysis final modeled SlMADS RIN (*Solanum lycopersicum* MADS Ripening inhibitor) structure (1-74 amino acid) as revealed by PROCHECK analysis

respectively which indicates better agreement. Thus, the overall validation analysis indicated that the main chain of SlMADS RIN protein (1-74 amino acids) was modeled with good confidence but side chains were modeled with much better confidence. Other verification programs also indicate the proper arrangement of residues around 3D space.

Table 3: Assessment of main chain parameter of modeled *SlMADS RIN** protein (1-74 amino acid) as revealed by PROCHECK analysis

			Comparison values			
Stereochemical parameter	No. of data pts	Parameter value	Typical value	Band width	No. of band widths from mean	
%-tage residues in A, B, L	48	87.5	83.8	10.0	00.4	Inside
Omega angle st dev	55	05.4	06.0	03.0	-0.2	Inside
Bad contacts / 100 residues	0	00.0	04.2	10.0	-0.4	Inside
Zeta angle st dev	70	01.1	03.1	01.6	-1.3	BETTER
H-bond energy st dev	37	00.8	00.8	00.2	-0.1	Inside
Overall G-factor	114	00.2	-0.4	00.3	01.8	BETTER

*SlMADS RIN: *Solanum lycopersicum* MADS Ripening inhibitor

Table 4: Assessment of side chain parameter of modeled SlMADS RIN* protein (1-74 amino acid) as revealed by PROCHECK analysis

			Comparison values			
Stereochemical parameter	No. of data pts	Parameter value	Typical value	Band width	No. of band widths from mean	
Chi-1 gauche minus st dev	3	6.4	18.1	6.5	-1.8	BETTER
Chi-1 trans st dev	22	5.3	19.0	5.3	-2.6	BETTER
Chi-1 gauche plus st dev	41	8.6	17.5	4.9	-1.8	BETTER
Chi-1 pooled st dev	66	8.5	18.2	4.8	-2.0	BETTER
Chi-2 trans st dev	33	9.0	20.4	5.0	-2.3	BETTER

*SlMADS RIN: *Solanum lycopersicum* MADS Ripening inhibitor

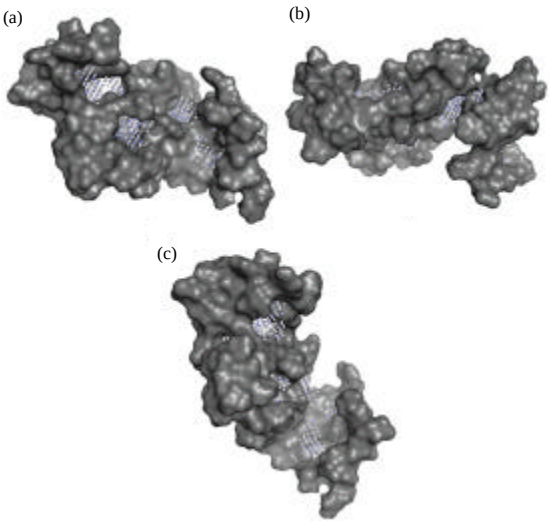


Fig. 10(a-c): Prediction of pocket on the modeled SlMADS RIN (*Solanum lycopersicum* MADS Ripening inhibitor) (1-74 amino acid) in different view, (a) Front view, (b) Side view and (c) Top view

Structure analysis: According to CATH analysis, the SlMADS RIN protein could be included into αβ2 layer sandwich class of protein. Surface of the modeled SlMADS RIN protein (1-74 amino acid) showed at least five active pockets (Fig. 10a-c). The surface area indicated hydrophobic residues were buried in the structure, whereas, hydrophilic and neutral residues were exposed on the surface (Fig. 11a, b).

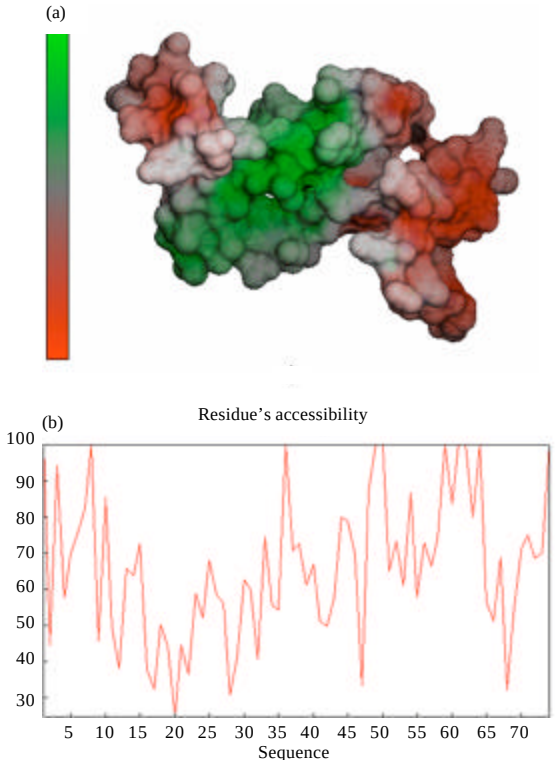


Fig. 11(a-b): Prediction of accessible surface area modeled SlMADS RIN (*Solanum lycopersicum* MADS Ripening inhibitor) protein (1-74 amino acid), (a) Representation in NOC 3.01 visualization software and (b) Graphical representation

In the present study the repeated occurrence of templates viz. 1TQE, 1EGW, 1N6J and 1C7U during searching with PSI BLAST and/or FUGUE, was suggestive of the suitability of these templates for predicting the 3D structures of SlMADS RIN protein. Each monomer of SlMADS RIN contains a long α-helix (αI) and two beta-strands (βI, βII) connected by a beta-turn forming the middle layer (β-hairpin). Together these two layers define the MADS-box. A second α-helix (αII) makes up the top layer or the major part of the MADS domain and packs against the beta-hairpin in an orientation diagonal. This observation was in great agreement with the MEF2A except the N-terminal extension (Santelli and Richmond, 2000).

CONCLUSION

As the structure of the plant MADS-box factor is not known, the present study is the first attempt in this direction known till date. Further the modeling of SlMADS RIN transcription factor was done in this study for the first time in literature. Thus, it can be concluded that this is the first report on prediction of three dimensional models for the plant MADS-box transcription factor proteins SlMADS RIN from tomato. From the evaluation data of PROCHECK, it was found that the predicted models were enough usual and good to represent the query proteins. Understanding three dimensional structures of SlMADS RIN help to understand the regulation of transcription factor. Based on the Template structure it is clearly observed that the theoretical structure generated is structurally similar to the template structure. This modeled structure can be used to predict the molecular function which received less attention in previous reports.

ACKNOWLEDGMENTS

The financial assistance from Department of Biotechnology, Govt. of India, through the grants no. BT/PR12989/Agr/16/231/2002 and BT/PR5278/Agr/16/468/2004 are gratefully acknowledged. The first author was a recipient of CSIR-NET JRF fellowship.

REFERENCES

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

Baker, D. and A. Sali, 2001. Protein structure prediction and structural genomics. Science, 294: 93-96.

Canutescu, A.A., A.A. Shelenkov and R.L. Dunbrack Jr., 2003. A graph theory algorithm for rapid protein side-chain prediction. Protein Sci., 12: 2001-2014.

Cara, B. and J.J. Giovannoni, 2008. Molecular biology of ethylene during tomato fruit development and maturation. Plant Sci., 175: 106-113.

Fiser, A., R.K. Do and A. Sali, 2000. Modeling of loops in protein structures. Protein Sci., 9: 1753-1773.

Garnier, J., J.F. Gibrat and B. Robson, 1996. GOR method for predicting protein secondary structure from amino acid sequence. Methods Enzymol., 266: 540-553.

Giovannoni, J.J., 2004. Genetic regulation of fruit development and ripening. Plant Cell, 16: S170-S180.

Guex, N. and M.C. Peitsch, 1997. SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. Electrophoresis, 18: 2714-2723.

Guleria, P. and S.K. Yadav, 2013. Insights into steviol glycoside biosynthesis pathway enzymes through structural homology modeling. Am. J. Biochem. Mol. Biol., 3: 1-19.

Hooda, V., P.B. Gundala and P. Chinthala, 2012. Sequence analysis and homology modeling of peroxidase from *Medicago sativa*. Bioinformation, 8: 974-979.

Ito, Y., M. Kitagawa, N. Ihashi, K. Yabe and J. Kimbara *et al.*, 2008. DNA-binding specificity, transcriptional activation potential and the rin mutation effect for the tomato fruit-ripening regulator RIN. Plant J., 55: 212-223.

Jamroz, M. and A. Kolinski, 2010. Modeling of loops in proteins: A multi-method approach. BMC Struct. Biol., Vol. 10. 10.1186/1472-6807-10-5

Joo, H., A.G. Chavan, R. Day, K.P. Lennox and P. Sukhanov *et al.*, 2011. Near-native protein loop sampling using nonparametric density estimation accommodating sparcity. PLoS Comput. Biol., Vol. 7. 10.1371/journal.pcbi.1002234

Laskowski, R.A., M.W. MacArthur, D.S. Moss and J.M. Thornton, 1993. PROCHECK: A program to check the stereochemical quality of protein structures. J. Applied Cryst., 26: 283-291.

Madhusudhan, M.S., M.A. Marti-Renom, R. Sanchez and A. Sali, 2006. Variable gap penalty for protein sequence-structure alignment. Protein Eng. Des. Sel., 19: 129-133.

Marsden, R.L. and C.A. Orengo, 2008. Target selection for structural genomics: An overview. Methods Mol. Biol., 426: 3-25.

Melo, F. and E. Feytmans, 1997. Novel knowledge-based mean force potential at atomic level. J. Mol. Biol., 267: 207-222.

Mo, Y., W. Ho, K. Johnston and R. Marmorstein, 2001. Crystal structure of a ternary SAP-1/SRF/c-fos SRE DNA complex. J. Mol. Biol., 314: 495-506.

Ohlson, T., B. Wallner and A. Elofsson, 2004. Profile-profile methods provide improved fold-recognition: A study of different profile-profile alignment methods. Proteins, 57: 188-197.

Orengo, C.A., A.D. Michie, S. Jones, D.T. Jones, M.B. Swindells and J.M. Thornton, 1997. CATH-α hierarchic classification of protein domain structures. Structure, 5: 1093-1109.

Pettersen, E.F., T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng and T.E. Ferrin, 2004. UCSF chimera-a visualization system for exploratory research and analysis. J. Comput. Chem., 25: 1605-1612.

Prajapat, R., A. Marwal, A. Sahu and R.K. Gaur, 2011. Phylogenetics and *in silico* docking studies between coat protein of Mimosa yellow vein virus and Whey a-lactalbumin. Am. J. Biochem. Mol. Biol., 1: 265-274.

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

Santelli, E. and T.J. Richmond, 2000. Crystal structure of MEF2A core bound to DNA at 1.5 A resolution. J. Mol. Biol., 297: 437-449.

Shi, J., T.L. Blundell and K. Mizuguchi, 2001. FUGUE: Sequence-structure homology recognition using environment-specific substitution tables and structure- dependent gap penalties. J. Mol. Biol., 310: 243-257.

Smith, A.A. and M.C. Plazas, 2011. *In silico* characterization and homology modeling of cyanobacterial phosphoenolpyruvate carboxylase enzymes with computational tools and bioinformatics servers. Am. J. Biochem. Mol. Biol., 1: 319-336.

Tan, S., Y. Hunziker, L. Pellegrini and T.J. Richmond, 2000. Crystallization of the yeast MATalpha2/MCM1/DNA ternary complex: General methods and principles for protein/DNA cocrystallization. J. Mol. Biol., 297: 947-959.

Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougi and D.G. Higgins, 1997. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res., 25: 4876-4882.

Vrebalov, J., D. Ruezinsky, V. Padmanabhan, R. White and D. Medrano *et al*., 2002. A MADS-box gene necessary for fruit ripening at the tomato ripening-inhibitor (rin) locus. Science, 296: 343-346.

Weisel, M., E. Proschak and G. Schneider, 2007. PocketPicker: Analysis of ligand binding-sites with shape descriptors. Chem. Cent. J., Vol. 1.