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

Sudip Kumar Sinha and Dibyendu N. Sengupta

1Department of Botany, Victoria Institution (College), 78B, A.P.C. Road, Kolkata-700009, India
2Division 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 SIMADS RIN (Solanum lycopersicum MADS Ripening inhibitor) seems to act as global regulator in climacteric fruit ripening process of tomato. Structure modeling of any plant MIKCC (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 SIMADS 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 SIMADS RIN protein was further validated by Ramachandran plot analysis using the PROCHECK tool. The submitted sequence of SIMADS 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 (-3968.569330). Ramachandran plot analysis revealed that 94.1% residues were in favored region and 05.9% residues were in allowed region. Result revealed that the SIMADS 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 SIMADS 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 SIMADS RIN and this modeled structure can be used to predict the molecular function of the protein.

Key words: Homology modeling, MADS-box, SIMADS 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 SIMADS RIN (Solanum lycopersicum MADS Ripening inhibitor) was expressed specifically in ripe tomato fruits and its expression is not significantly influenced by ethylene. But the SIMADS 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, SIMADS 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 SIMADS RIN protein.

MATERIALS AND METHODS

Domain identification: Domain identification of SIMADS RIN protein was carried out by submitting the SIMADS 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 SIMADS 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 SIMADS 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 SIMADS 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 SIMADS RIN protein identified by CDD of NCBI, fragment of SIMADS 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 SIMADS 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 SIMADS 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 SIMADS RIN protein 1-74 amino acid. Among the five modeled structures that were generated for SIMADS 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 SIMADS 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 Cα-backbone for all the loops of SIMADS 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 Peytmans, 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 SIMADS RIN protein was done with the help of SWISS-pdb Viewer (Guex and Peitsch, 1997), using both Steepest Descent and Conjugate Gradients of AMBER97 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 SIMADS 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 SIMADS RIN protein into Class, Fold etc.

Active site identification: Accessible surface area of modeled SIMADS 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 SIMADS 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 SIMADS RIN protein, namely MADS_MEF2 like and K-box super family (Fig. 1). The submitted sequence of SIMADS RIN protein to PSI BLAST tool, for retrieving the homologous sequences deposited in PDB 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

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

**Fig. 1:** Schematic diagram to show different PDB (Protein Data Bank) hits in respect to SIMADS RIN (Solanum lycopersicum MADS Ripening inhibitor) amino acid sequence as retrieved after seventh iteration of PSI (Position Specific Iterative)-BLAST search.

![Color key for alignment scores](image)
Table 1: Summary of the best template sequence profile that was generated at the end of seventh iteration of PSI BLAST* analysis using SIMADS RIN*

<table>
<thead>
<tr>
<th>SIMADS RIN region</th>
<th>PDB* ID of Template</th>
<th>Region of template sequence aligned</th>
<th>Identity (%)</th>
<th>Gap (%)</th>
<th>e-value</th>
<th>Annotation of template</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-74</td>
<td>EE7E chain F, Q, R and S</td>
<td>1-74</td>
<td>54</td>
<td>1</td>
<td>1e-29</td>
<td>Myocyte Enhancer Factor-2</td>
</tr>
<tr>
<td>1EOW chain A, B, C, and D</td>
<td>2-74</td>
<td>57</td>
<td>1</td>
<td>1e-29</td>
<td>Myocyte Enhancer Factor-2</td>
<td></td>
</tr>
<tr>
<td>1N6J chain A &amp; B</td>
<td>2-74</td>
<td>53</td>
<td>1</td>
<td>4e-29</td>
<td>Myocyte Enhancer Factor-2</td>
<td></td>
</tr>
<tr>
<td>1C7U chain A and B</td>
<td>2-74</td>
<td>57</td>
<td>1</td>
<td>1e-27</td>
<td>Myocyte Enhancer Factor-2</td>
<td></td>
</tr>
<tr>
<td>1K6O chain B and C</td>
<td>2-59</td>
<td>43</td>
<td>0</td>
<td>3e-21</td>
<td>Serum responsive factor</td>
<td></td>
</tr>
<tr>
<td>1MN8 chain A and B</td>
<td>3-59</td>
<td>47</td>
<td>0</td>
<td>1e-20</td>
<td>MCM1</td>
<td></td>
</tr>
</tbody>
</table>

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

1MN8 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, >29% 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 Ca 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, 1MN8, 1K60 and 1SRS MADS transcription factor entries are present. Most Arabidopsis MADS-box transcription factors only have obvious similarity with 1MN8, 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 SIMADS 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 SIMADS 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 SIMADS RIN protein that was aligned with various template sequences of PSI-BLAST, 1TQH, 1EGW, 1N6J and 1C7U were selected as the template sequences.
Table 2: Summary of structural comparison by FUGUE program for SIMADS RIN* protein (1-74 amino acid)

<table>
<thead>
<tr>
<th>Profile Hit</th>
<th>PLEN*</th>
<th>RAW8*</th>
<th>RVN*</th>
<th>Z_score*</th>
<th>ZORI*</th>
<th>AL*</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRF-TF</td>
<td>85</td>
<td>47</td>
<td>193</td>
<td>19.40</td>
<td>19.87</td>
<td>02</td>
</tr>
<tr>
<td>hs1356a</td>
<td>77</td>
<td>-68</td>
<td>50</td>
<td>4.64</td>
<td>4.49</td>
<td>02</td>
</tr>
<tr>
<td>hs13dia1</td>
<td>166</td>
<td>-51</td>
<td>101</td>
<td>4.12</td>
<td>4.87</td>
<td>00</td>
</tr>
<tr>
<td>hs13ae2a</td>
<td>20</td>
<td>-190</td>
<td>17</td>
<td>4.07</td>
<td>3.92</td>
<td>02</td>
</tr>
<tr>
<td>hs13m01a</td>
<td>78</td>
<td>-128</td>
<td>13</td>
<td>3.97</td>
<td>4.74</td>
<td>02</td>
</tr>
<tr>
<td>hs13m09a</td>
<td>27</td>
<td>46</td>
<td>15</td>
<td>3.78</td>
<td>3.68</td>
<td>22</td>
</tr>
<tr>
<td>hs13p01</td>
<td>32</td>
<td>41</td>
<td>16</td>
<td>3.65</td>
<td>3.50</td>
<td>22</td>
</tr>
<tr>
<td>hs13m03a</td>
<td>318</td>
<td>-273</td>
<td>109</td>
<td>3.48</td>
<td>4.84</td>
<td>00</td>
</tr>
<tr>
<td>hs13m04a</td>
<td>78</td>
<td>-231</td>
<td>60</td>
<td>3.41</td>
<td>4.69</td>
<td>00</td>
</tr>
<tr>
<td>hs13m07a</td>
<td>185</td>
<td>-68</td>
<td>83</td>
<td>3.39</td>
<td>4.44</td>
<td>00</td>
</tr>
</tbody>
</table>

*SIMADS RIN: Solanum lycopersicums MADS Ripening inhibitor; PLEN: Profile Length, RAW8: Raw alignment scores, 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 SIMADS 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 SIMADS 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 SIMADS RIN protein (Fig. 6a). Moreover, from GOR4 server prediction it was found that, SIMADS 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 Ca-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 SIMADS RIN (Solanum lycopersicum MADS Ripening inhibitor) (1-74 amino acid) with the templates using MODELLER program using ‘align2d_multi’ command, MADS: Minichromosome maintenance-1, Agamous, Deficiens and Serum response factor.

Fig. 5: Initial backbone model of SIMADS RIN (Solanum lycopersicum MADS Ripening inhibitor) (1-74 amino acid) as build by ‘model_multi’ of MODELLER. (a) Ca-backbone model coloured by secondary structure type, (b) Backbone represented by ribbon and coloured by secondary structure type

(from-5871.276337 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 SIMADS RIN were modeled with RMSD of 0.233 Å. The loop region 50-66 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 SIMADS 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
Summary of GOR4 prediction:
- Alpha helix (Hh) : 131 is 54.13%
- 3_1 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 coil (Cc) : 84 is 34.71%
- Ambiguous states (?) : 0 is 0.00%
- Other states : 0 is 0.00%

Fig. 5(a-b): Secondary structure prediction of SIMADS 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 SIMADS RIN [c = random coil; h = alpha helix; e = extended strand], (b) Graphical and tabular representation of secondary structural elements present in SIMADS RIN.

Fig. 7: Comparison of DOPE (Discrete Optimized Protein Energy) value for entire chain of modeled SIMADS 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 SIMADS RIN (Solanum lycopersicum MADS Ripening inhibitor) in respect to target PDB (Protein 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 SIMADS 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 SIMADS RIN protein (1-74 amino acid). Structural superimposition of Cα-backbone of modeled SIMADS 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 Hoods et al. (2012). Ramachandran plot analysis of modeled SIMADS RIN protein (1-74 amino acid) revealed that 94.1% residues were in favoured region and 5.9% 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 SIMADS 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,
Fig. 9: Ramachandran plot analysis final modeled SIMADS 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 SIMADS 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 SIMADS RIN* protein (1-74 amino acid) as revealed by PROCHECK analysis

<table>
<thead>
<tr>
<th>Stereocchemical parameter</th>
<th>No. of data pts</th>
<th>Parameter value</th>
<th>Typical value</th>
<th>Band width</th>
<th>widths from mean</th>
<th>No. of band</th>
</tr>
</thead>
<tbody>
<tr>
<td>%a-tage residues in A, B, L</td>
<td>48</td>
<td>87.5</td>
<td>83.8</td>
<td>10.0</td>
<td>0.04</td>
<td>Inside</td>
</tr>
<tr>
<td>Omega angle st dev</td>
<td>65</td>
<td>00.4</td>
<td>06.0</td>
<td>06.0</td>
<td>-0.2</td>
<td>Inside</td>
</tr>
<tr>
<td>Bad contacts / 100 residues</td>
<td>0</td>
<td>00.0</td>
<td>04.2</td>
<td>10.0</td>
<td>-0.4</td>
<td>Inside</td>
</tr>
<tr>
<td>Zeta angle st dev</td>
<td>70</td>
<td>01.1</td>
<td>06.1</td>
<td>01.6</td>
<td>-1.3</td>
<td>BETTER</td>
</tr>
<tr>
<td>H-bond energy st dev</td>
<td>37</td>
<td>00.8</td>
<td>00.8</td>
<td>00.2</td>
<td>-0.1</td>
<td>Inside</td>
</tr>
<tr>
<td>Overall G-factor</td>
<td>114</td>
<td>00.2</td>
<td>-0.4</td>
<td>00.3</td>
<td>01.8</td>
<td>BETTER</td>
</tr>
</tbody>
</table>

*SIMADS RIN: Solanum lycopersicum MADS Ripening inhibitor

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

<table>
<thead>
<tr>
<th>Stereocchemical parameter</th>
<th>No. of data pts</th>
<th>Parameter value</th>
<th>Typical value</th>
<th>Band width</th>
<th>widths from mean</th>
<th>No. of band</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-1 gauche minus st dev</td>
<td>3</td>
<td>6.4</td>
<td>18.1</td>
<td>6.5</td>
<td>-1.8</td>
<td>BETTER</td>
</tr>
<tr>
<td>Chi-1 trans st dev</td>
<td>22</td>
<td>5.3</td>
<td>19.9</td>
<td>5.3</td>
<td>-2.6</td>
<td>BETTER</td>
</tr>
<tr>
<td>Chi-1 gauche plus st dev</td>
<td>41</td>
<td>8.6</td>
<td>17.5</td>
<td>4.9</td>
<td>-1.8</td>
<td>BETTER</td>
</tr>
<tr>
<td>Chi-1 potted st dev</td>
<td>65</td>
<td>8.5</td>
<td>18.2</td>
<td>4.8</td>
<td>-2.0</td>
<td>BETTER</td>
</tr>
<tr>
<td>Chi-2 trans st dev</td>
<td>33</td>
<td>9.0</td>
<td>20.4</td>
<td>5.0</td>
<td>-2.3</td>
<td>BETTER</td>
</tr>
</tbody>
</table>

*SIMADS RIN: Solanum lycopersicum MADS Ripening inhibitor

Fig. 10(a-c): Prediction of pocket on the modeled SIMADS 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 SIMADS RIN protein could be included into aβ2 layer sandwich class of protein. Surface of the modeled SIMADS 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 SIMADS 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 SIMADS RIN protein. Each monomer of SIMADS RIN contains a long $\alpha$-helix ($\alpha$I) and two beta-strands ($\beta$I, $\beta$II) connected by a beta-turn forming the middle layer ($\beta$-hairpin). Together these two layers define the MADS-box. A second $\alpha$-helix ($\alpha$II) 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 SIMADS 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 SIMADS 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 SIMADS 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/16/458/2004 are gratefully acknowledged. The first author was a recipient of CSIR-NET JRF fellowship.

REFERENCES


