In silico prediction and characterization of three-dimensional structure of Actin-1 of *Arabidopsis thaliana*

Mousumi Sahu, Budheswar Dehury, Ranjan Sarmah, Smita Sahoo, Jagajjit Sahu, Kishore Sarma, Priyabrata Sen, Mahendra Kumar Modi, Madhumita Barooah*

Department of Agricultural Biotechnology, Assam Agricultural University, Assam, India

*Corresponding author: m17barooah@yahoo.co.in

Abstract

Actin-1 is a ubiquitous protein belonging to the reproductive class of Actin family in *Arabidopsis thaliana*. This protein is involved in the formation of filaments that are major components of the cytoskeleton. Despite the importance of this protein, very little information is available regarding its structure and function in plants. In this study, analysis of the protein sequence was done and comparative model of Actin-1 was constructed (UNIPROT ID: P0CJ46) from *Arabidopsis thaliana* using the crystal structure of *Dictyostelium discoideum* actin (PDB ID: 1NLV-A) as template employing Modeller version 9.9. The stable structure was generated by 5 nanosecond molecular dynamics simulation steps using GROMOS43A1 96 force field that characterized its structural and dynamic feature. The biochemical function of the final simulated structure was also investigated using PROFUNC. The molecular simulation study suggested that the modeled Actin-1 protein retain its stable conformation in aqueous solution. The predicted binding sites in the modeled Actin-1 protein are very informative for further protein-ligand interaction study.

Keywords: Actin-1, *Arabidopsis thaliana*, comparative modeling, molecular dynamics simulation

Introduction

Arabidopsis thaliana is unique among plant model organisms and contains two major actin gene classes: a vegetative class that is expressed predominantly in leaves, stems, roots, petals, and sepals; and a reproductive class that is strongly expressed in pollens, ovules, and embryonic tissues (McDowell et al., 1996; Kandasamy, McKinney and Meagher, 2002). *ACT1* gene, belonging to reproductive class (McKinney et al., 1995), codes for Actin-1 protein that is involved in the formation of filaments, major components of the cell cytoskeleton. These filaments interact with myosin to produce a sliding effect, which is the basis of the muscular contraction and many aspects of cell motility, including cytokinesis. The actin cytoskeleton controls polar cell growth through its interaction with several actin binding proteins, such as actin depolarizing factor (Dong et al., 2001), profilin (Chen et al., 2002; Clarke et al., 1998), Rho family GTPase (Opin et al., 1998; Cheung et al., 2002; Fu et al., 2002), and the calcium signaling pathway (Li et al., 2005; Franklin-Tong, 1999; Li et al., 1999). Effective regulation of actin turnover by actin regulators may be critical for pol-

len tube growth (Chen et al., 2002) and for polar cell expansion in cell types other than root hair and trichome (Fu et al., 2002).

A more complete understanding of the regulation of actin dynamics will be required before the mechanism by which actin-dependent cellular processes occur can be fully understood. *In silico* study of this protein family can provide an insight and a basis for understanding its structure and functional aspect.

Due to technical difficulties, there is a vast gap in the pace of protein sequence generation and experimental determination of the three dimensional (3D) structure. In the absence of the crystal structure, homology modeling provides a way to obtain a structural insight into proteins. In this paper, structure of Actin-1 protein of *A. thaliana* was generated employing a comparative modeling approach. The models were subjected to molecular dynamics (MD) simulations to study the time evolution as well as time averaged values of structural properties. The predicted structure can serve as a model for various other experimental purposes in Arabidopsis as well as in several other related species.

Materials and methods

Sequence analysis

The reviewed amino acid sequence of Actin-1 of Arabidopsis (target) was retrieved from the UniProt (http: www.uniprot.org/) sequence database (UniProt ID: P0CJ46). Domains and motifs were predicted using InterProScan (Quevillon et al., 2005) and Multiple Em for Motif Elicitation (MEME) (Bailey and Elkan, 1994), respectively, using default parameters to gain an insight into the function of Arabidopsis Actin-1 protein. Protein disorder in Actin-1 protein of Arabidopsis was predicted using the DisEMBL (Linding et al., 2003) server.

3D structure generation

Our initial search in the Protein Data Bank (PDB) for 3D structure of Actin-1 protein of *A. thaliana* was unsuccessful. Therefore, we started the exercise of developing a 3D model of Actin-1 of *A. thaliana*. A basic local alignment search (BLASTP) (Altshul et al., 1990) search was performed against PDB (Bermen et al., 2000) with default parameters to find suitable templates for homology modeling of Actin-1 protein (P0CJ46). Based on the maximum identity, query coverage and lower E-value, the crystal structure of *Dictyostelium discoideum,* actin was solved at 1.8Å resolution (PDB ID: 1NLV-A) (Vorobiev et al., 2003) was selected as template. Actin-1 showed a sequence identity of 90% and query coverage of 99% with the selected template 1NLV. The sequence alignment of Actin-1 (query) and the template (1NLV-A chain) was carried out using the CLUSTALW (Thompson, Higgins and Gibson, 1994) program.

The academic version of MODELLER version 9.9 (http://www.salilab.org/modeler) was used for 3D structure generation based on the information obtained from sequence alignment. The MODELLER software employs probability density functions (pdfs) rather than energy as the spatial restraints (Sali and Blundell, 1993a, 1993b; Sali and Overington, 1994). The 3D model of a protein is obtained by optimization of the molecular pdf, so that the model violates the input restraints as little as possible. The molecular pdf was derived as a combination of pdfs restraining individual spatial features of the whole molecule (Sali et al., 1995). Out of 20 models generated by MODELLER, the one with the best G-score of PROCHECK (Laskowski et al., 1993) and with the best VERIFY3D (Eisenberg, Luthy and Bowie, 1997) profile

was subjected to energy minimization. The Swiss-Pdb Viewer (Guex and Peitsch, 1997) was implemented for energy minimization; using a harmonic constraint of 100 kJ mol^{-1} with the steepest descent and conjugate gradient techniques using GROMOS energy minimization parameters. During these steps, the quality of the initial model was improved. VERIFY3D (a structure evaluation server) was used to check the residue profiles of the obtained 3D models. In order to assess the stereochemical qualities of the 3D models, the PROCHECK analysis was performed. Quality evaluation of the model for the environment profile was also predicted using ERRAT (structure evaluation server) (Colovos and Yeates, 1993). The final refined model was evaluated for its atomic contacts using the What_check program (Hooft et al., 1996) to identify bad packing of side chain atoms or unusual residue contacts. ProQserver program (Cristobal et al., 2001) was also employed that predicts quality of a protein model using a neural network approach, considering the structural feature of the theoretical model.

MD simulation study

MD simulations were conducted for the modeled systems in an explicit solvent using the GROMACS (Groningen Machine for Chemical Simulations) 4.5.3 package (Hess et al., 2008), the GROMOS 96 43A1 force field (Scott et al., 1999) and the flexible SPC water model. The initial structure was immersed in a periodic water box of cubic shape (0.9 nm thick). The system was neutralized with twelve Na^{+} counterions. The solvated system was then subjected to further energy minimization (maximum number of steps: 1000) to remove steric conflicts between the protein and water molecules, using the steepest descent integrator. Convergence was achieved in energy minimization when the maximum force was less than 1000 kJ mol^{-1} nm^{-1}. The energy-minimized model was subjected to position-restrained MD under NPT conditions, keeping the number of particles (N), the system pressure (P) and the temperature (T) constant. This was carried out for 50,000 steps for a total of 100 ps. LINCS (Hess et al., 1997) constraints were performed for all bonds, keeping the whole protein molecule fixed and allowing only the water molecule to move to equilibrate with respect to the protein structure. The reference temperature for coupling (*via* v-rescale temperature coupling) was 300 K, and a pressure of 1 atm was

Table 1. Location of domains predicted using InterProScan

Method	Description	Matches
PRINTS	ACTIN	28-37, 51-62, 63-85, 117-130, 142-61, 238-254
PANTHER	FAMILY NOT NAMED	2-377
PFAM	Actin	5-377
SMART	no description	7-377
PROSITE	ACTINS_1	55-65
PROSITE	ACTINS_2	358-366
PROSITE	ACTINS_ACT_LIKE	106-118
GENE3D	no description	37-70
GENE3D	no description	271-350
GENE3D	no description	2-377
PANTHER	ACTIN	2-377
SUPERFAMILY	Actin-like ATPase domain	149-377, 5-155

Table 2. Motifs predicted using MEME

| Motif	Width	Sites	E-value	Start	P-value	Domain sites
Motif 1	14	2	1.2e-001	79	8.32e-19	KYPIEHGIVN***NWDDMEKIWHHTFY*** NELRVAPEEH
				116	2.87e-16	LTEAPLNPKA***NREKMTQIMFETFN***APAMYVAIQA
Motif 2	9	2	1.4e+001	368	8.03e-13	IAKAEYDESG ***PSIVHRKCF***
				33	3.81e-11	AGDDAPRAVF***PSIVGRPRH*** TGVMVGMGQK
Motif 3	6	2	3.3e+001	352	2.21e-08	IGGSILASLS ***TFQQMW*** IAKAEYDESG
				255	3.12e-08	DGQVITIGSE***RFRCPE*** VLYQPSMIGM
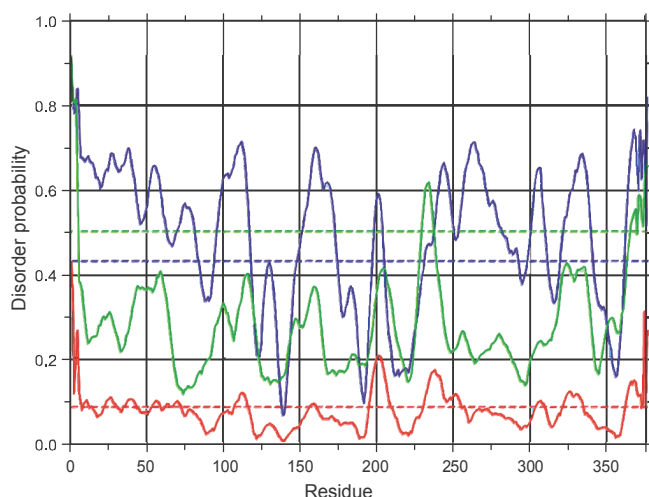
Fig. 1. Disordered regions in Actin-1 protein predicted by DisEMBL. The green, red, and blue lines represent the predictions for missing coordinates, the hot loop network and the coil, respectively. The horizontal lines correspond to the random expectation level for each predictor

maintained by the Parrinello-Rahman algorithm. The final MD calculations of 500,000 steps were carried out for 1 ns using the particle mesh Ewald (PME) electrostatics method under NPT conditions. The results were analyzed using the standard software provided in the GROMACS package.

Function characterization

The biochemical function of the final simulated structure of Actin-1 protein of Arabidopsis was predicted using ProFunc (Laskowski, Watson and Thornton, 2005a, 2005b). The functionally active regions, possible nests or enzyme catalytic sites were explored in the modeled Actin-1 protein using ProFunc server (Watson and Milner-White, 2002a; Watson and Milner-White, 2002b; Pal et al., 2002; Laskowski, Watson and Thornton, 2005a; Porter, Bartlett and Thornton, 2004). Sequences showing significant sequence similarity (cut off

```
Actin-1   MADGEDIQPLVCDNGTGMVKAGFAGDDAPRAVFPSIVGRPRHTGVMVGMGQKDAYVGDEA 60
1NLV_A    --DGEDVQALVIDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHTGVMVGMGQKDSYVGDEA 58
          ****:*.** ***:** *****************************:***:******

Actin-1   QSKRGILTLKYPIEHGIVNNWDDMEKIWHHTFYNELRVAPEEHPILLTEAPLNPKANREK 120
1NLV_A    QSKRGILTLKYPIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREK 118
          ******************.*.***************************:***********

Actin-1   MTQIMFETFNAPAMYVAIQAVLSLYASGRTTGIVLDSGDGVSHTVPIYEGYALPHAILRL 180
1NLV_A    MTQIMFETFNTPAMYVAIQAVLSLYASGRTTGIVMDSGDGVSHTVPIYEGYALPHAILRL 178
          **********:***********************:*************************

Actin-1   DLAGRDLTDALMKILTERGYSFTTTAEREIVRDIKEKLCYIALDYEQELETAKTSSSVEK 240
1NLV_A    DLAGRDLTDYMMKILTERGYSFTTTAEREIVRDIKEKLAYVALDFEQEMATAASSSALEK 238
          *********  :***********************  :  ***:***: ** :**:*:**

Actin-1   NYELPDGQVITIGSERFRCPEVLYQPSMIGMENAGIHETTYNSIMKCDVDIRKDLYGNIV 300
1NLV_A    SYELPDGQVITIGNERFRCPEALFQPSFLGMESAGIHETTYNSIMKCDVDIRKDLYGNVV 298
          .***********.*******.*::**::***:*************************:*

Actin-1   LSGGTTMFPGIADRMSKEITALAPSSMKIKVVAPPERKYSVWIGGSILASLSTFQQMWIA 360
1NLV_A    LSGGTTMFPGIADRMNKELTALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWIS 358
          ***************.**:******:*****:**************************:

Actin-1   KAEYDESGPSIVHRKCF 377
1NLV_A    KEEYDESGPSIVHRKCF 375
          * ***************
```

Fig. 2. Sequence Alignment of template (1NLV_A) and Actin-1 protein from *Arabidopsis thaliana.* The identical (asterisk) or similar (dot or colon) amino acids are also marked

>85%) with Actin-1 sequence of Arabidopsis from the UniProtKB database were retrieved for further analysis. The SSM (Secondary Structure Matching) (Krissinel and Henrick, 2004) program of ProFunc identified proteins in the PDB that had a similar fold as the modeled Actin-1 protein structure.

Phylogenetic studies

Arabidopsis thaliana contains ten sub-classes of Actin proteins. To classify the reproductive and vegetative class of actin family of Arabidopsis, the actin proteins of this plant with the following UniProt IDs: P0CJ46 (Actin-1), P0CJ47 (Actin-3), P53494 (Actin-4), P53497 (Actin-12), P53496 (Actin-11), P53492 (Actin-7), Q96292 (Actin-2), Q96293 (Actin-8), Q8RYC2 (Actin-5), P93738 (Actin-9) were retrieved from the UniProt sequence database. They were classified and analyzed by applying 1000 iteration bootstrapping steps employed in MEGA.

Result and discussion

Sequence analysis

Identification of certain domains and motifs may help in defining a functional profile of the protein. Table 1 shows the location of predicted domains. It indicates that the modeled Actin-1 protein structure of Arabidopsis possesses two ATPase domains involving residues 149-377 and 5-155. Details about the location and occurrence of the motifs calculated using MEME server are summarized in Table 2. Disordered regions in proteins

facilitate their interactions with other proteins and thus allow more modification sites in the modeled Actin-1 protein structure. In this study, disordered regions were predicted using the DISEMBL server with disorder probability thresholds for loop/coil (0.516), hot loop (0.1204) and Remark-465 (0.6). The intrinsic disorder profiles of the obtained Actin-1 protein are illustrated in Figure 1. The disorder peaks were observed at 0.786 for loop/coil, at 0.43 for hot loop and at 0.914 for Remark-465 (Fig. 1). It is noteworthy that the amino acid residues ranging from 230-237 are displayed as disordered by Remark-465 (i.e. the residues that are present in the SEQRES records but are not present in the coordinates section). The amino acid residues in regions 1-84, 95-117, 149-174, 232-292, 299-311, 320-340, 362-377 are disordered as predicted by loops/coils. Similarly, the residues in the regions of 196-209, 231-255, 322-334, and 363-377 are disordered by hot loop definition with a high degree of mobility as determined from C-alpha temperature (B) factors.

Model building

Employing homology modeling protocol, the structure of Actin-1 of *A. thaliana* was constructed using the crystal structure of 1NLV_A as template. The sequence of 1NLV_A showed 90% identity with the query sequence of Actin-1 of Arabidopsis in the BLASTP search performed against PDB with default parameters to find suitable templates for homology modeling. The E-value and query coverage, between the target (Arabidopsis Actin-1) and the template (1NLV_A), were 0.0 and 99%, respectively. The alignment between the target and the query sequence was analyzed using ClustalW program (Fig. 2). The modeling was carried out and was followed by a rigorous refinement of the model by means of energy minimization using GROMOS force field. The final stable structure of Actin-1 protein of Arabidopsis protein achieved using Discovery Studio version 3.5 (DS3.5; Accelrys Inc. San Diego, CA, USA) is shown in Figure 3.

Protein structure validation

The stereochemical quality of the homology model of Arabidopsis Actin-1 protein was analyzed using PROCHECK. It was found that the phi/psi angles of 95.4% residues fell in the most favored regions and 4.6% in the allowed region of the Ramachandran plot. The overall PROCHECK G-factor for the homology modeled struc-
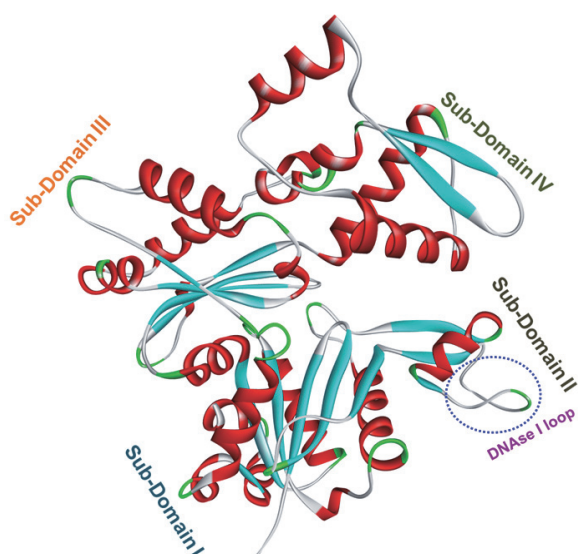
Fig. 3. Ribbon representation of Actin-1 protein of Arabidopsis. The *α*-helices and *β*-sheets (parallel) are shown as helices and ribbons, respectively. The rest is shown as loops. The figure was prepared using Discovery Studio 3.5 (DS3.5; Accelrys Inc. San Diego, CA, USA)

ture was 0.13. The PROCHECK result validated the homology modeled structure (Table 3). The model was also validated using ERRAT graph (Fig. 4). The ERRAT quality factor of about 91.6% suggests that the model has good quality as a score higher than 50 is acceptable for a reasonable model. The energetic architecture as predicted by PROSA (Sippl et al., 1993; Wiederstein and Sippl, 2007) score was negative (−9.94) for the modeled protein. The value was quite proximal to that of the template (−9.51), which indicates the reliability of the model. Its high quality has also been confirmed byVERIFY3D server as 96.56% of residues of the modeled protein showed a score higher than 0.2. The results of a WHAT_CHECK also indicated the correctness of the modeled structure

(Table 4). Moreover, the LGscore and MaxSub of the theoretical model were calculated by ProQ server to be 5.818 and 0.252, respectively, which suggests that the quality of the modeled protein structure is high. The final model was submitted to the Protein Model Data Base (http://mi.caspur.it/PMDB/) with the PMDB id: PM0077749.

The structural topology of the theoretical structure is very similar to the template structure. Moreover, like mammal actin, Arabidopsis Actin-1 also possesses two major domains (I and II) (Fig. 3). These two domains can be further divided into smaller sub-domains, one consisting of sub-domain I and sub-domain II, and the other domain being composed of sub-domain III and sub-domain IV (Fig. 3). Actin-1 of Arabidopsis shows a high degree of structural similarity with vertebrate and non-vertebrate actins. However, the DNase I loop in the sub-domain II (residues: 33-69) is considered to be the most variable region and has been highlighted in Figure 3.

MD simulation study

The MD simulation of the modeled Actin-1 protein was performed and the resulting trajectory was analyzed to study the motional properties of the protein. The time evolution of the root mean square deviation (RMSD) was computed for the modeled structure of the protein by taking the whole protein as the initial structure (Fig. 5). The steepest descent method of energy minimization for the solvated protein model showed that the maximum force dropped below the defined value after 423 steps. It is evident from Figure 4 that the RMSD increased with the highest pickup to 580 ps, then decreased down to 656 ps and finally reached the equilibrium. Based on the intrinsic dynamics, the structural stability and the
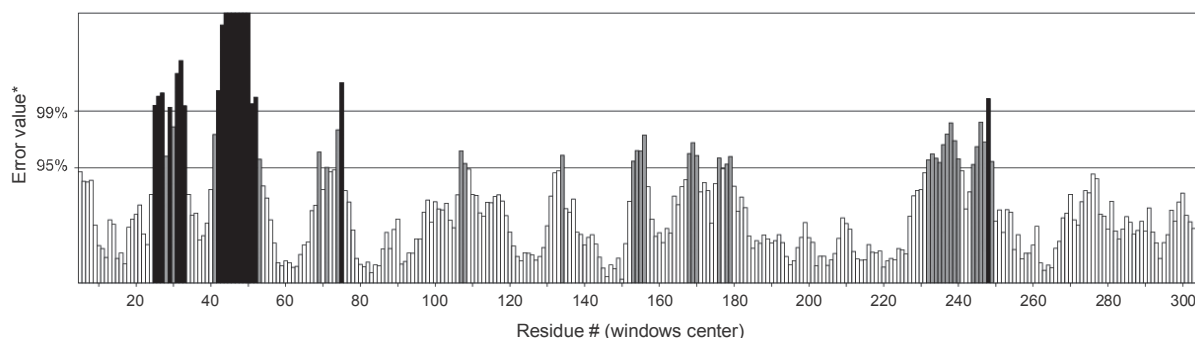

Fig. 4. 3-D profiles of Actin-1 protein of Arabidopsis, verified using ERRAT server. The overall quality score indicates that residues are reasonably folded

Table 3. Quality of the main chain and side chain parameters of modeled Actin-1 protein providing
a comparative account of values observed for modeled structure (parameter value)
and values obtained for well refined structures at the same resolution (typical values)

Chain	Stereo-chemical parameter	No. of data pts	Parameter value	Typical value	Band width	No. of band width from mean parameter
Main Chain Parameter	percentage of residues in A, B, L	325	95.4	83.8	10.0	1.2
	omega angle stdev	376	4.6	6.0	3.0	−0.5
	bad contacts / 100 residues	2	0.5	4.2	10.0	−0.4
	zeta angle stdev	347	1.7	3.1	1.6	−0.9
	H-bond energy stdev	248	0.8	0.8	0.2	0.0
	overall G-factor	377	0.1	−0.4	0.3	1.7
Side Chain Parameter	chi-1 gauche minus stdev	54	7.8	18.1	6.5	−1.6
	chi-1 trans stdev	99	7.3	19.0	5.3	−2.2
	chi-1 gauche plus stdev	144	8.6	17.5	4.9	−1.8
	chi-1 pooled stdev	297	8.9	18.2	4.8	−1.9
	chi-2 trans stdev	99	11.3	20.4	5.0	−1.8

Table 4. Quality indicators as calculated from WHAT_CHECK

Structure	Z-score	RMS-Z	Score
2nd generation packing quality	−1.406	bond lengths	0.609
Ramachandran plot appearance	0.766	bond angles	1.022
Chi-1/chi-2 rotamer normality	−0.325	Omega angle restraints	0.830
Backbone conformation	−4.299	side chain planarity	1.759
		improper dihedral distribution	1.838
		inside/outside distribution	0.989
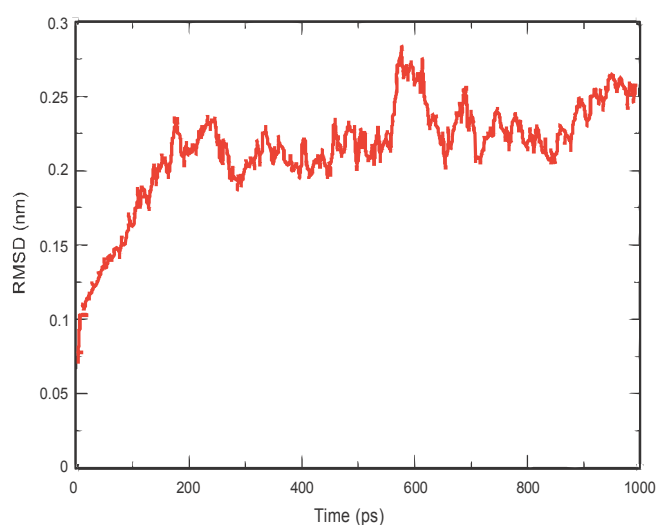
Fig. 5. Trajectories of the overall backbone (RMSD) of Actin-1 protein structure of *Arabidopsis thaliana*. The x-axis represents the simulation time in picoseconds. The y-axis represents RMSD in nm unit

improved relaxation of the modeled protein, the energy of the structure was calculated (Fig. 6). The energy and RMSD calculations for Actin-1 protein demonstrated that the protein was not very flexible over the timescale of the MD simulations.

The simulated structure was aligned with the template structure (1NLV_A) (Fig. 7) using SuperPose version 1.0 (Maiti et al., 2004). The structural superimposition of Cα trace of the target model after MD simulation over template structure 1NLV_A (Fig. 4) resulted in an RMSD of 2.71Å using SuperPose. The identity and similarity between the modeled structure and template structure were 87.0% and 93.4%, respectively. Summarizing the results presented above, it can be concluded that modeled Actin-1 is a reliable model for subsequent modeling exercises. In addition, the structural comparison between the target and template model showed a high de-
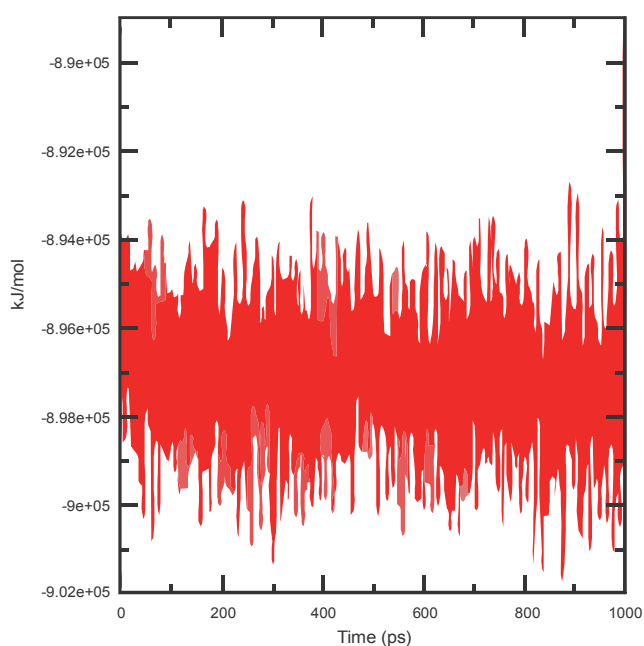
Fig. 6. Calculated energy *vs.* time plot using GROMACS software. The x-axis represents the simulation time in picoseconds. The y-axis represents energy in nm unit
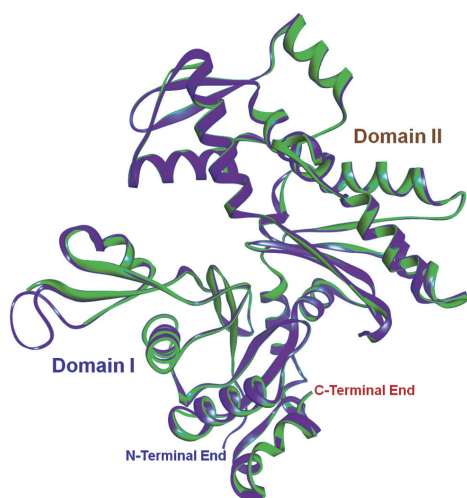

Fig. 7. Superposition of the final simulated Actin-1 protein (Blue) of Arabidopsis and *1NLV_A* (Green) predicted by SuperPose v1.0

gree of similarity for both domains (domain I and domain II) (see Fig. 7).

Function characterization

Nests are structural motifs that are often found in functionally important regions of protein structures. Seven nests were explored in the modeled structure of Actin-1 protein by ProFunc Nest analysis (Table 5). The nest score indicates how functionally significant the nest is likely to be. The score takes into account the number of the NH atoms that are accessible to the solvent, the conservation score of their parent residues, and whether they are associated with one of the larger surface clefts. A score of 2.0 suggests the nest being a functionally significant one. The modeled structure of Actin-1 protein of Arabidopsis possesses four functionally active nests, which in the present study, were signified by their nest score equal to 2.0. The residues of each nest that forms the RIGHT and LEFT region of the Ramachandran plot and that of NH which makes part of the nest have been listed in Table 5. The residue conservation score for each nest was determined from a multiple sequence alignments of the protein's sequence against BLAST hits from the UniProt sequence database. It was found that the residues were well conserved in all the predicted nests having residue conservation score 1.0 except for Ala 131 that had the conservation score value of 0.63 (i.e., the value is less than 2).

The ProFunc result revealed that the modeled Actin-1 protein structure has an enzymatically active site. The resulted similarity score between the neighborhood around the matched side chains of the modeled Actin-1 structure and side chains of the parent structure (PDBID: 2PHK) is about 130. The percentage of the structural similarity and E-value between Actin-1 structure and 2PHK are 70.4% and about 0.7, respectively. This above-mentioned structural similarity indicates that the query and target protein structures can be structurally superposed. The template structure represents the crystal structure of a phosphorylase kinase. It indicates a possible enzymatic property of the modeled Actin-1 structure having an enzymatically active template. The predicted secondary structural topology of the Arabidopsis Actin-1 is represented in Figure 8.

Using SSM (Secondary Structure Matching) program, protein structures in the PDB with secondary structure topology similar to that of the Actin-1 protein structure, were identified. The SSM program superimposed the modeled Actin-1 structure with the PDB structure on the basis of the matched secondary structure elements (SSEs) such as helices and strands between the two structures. The secondary structure architecture of the modeled protein structure of Actin-1 of Arabidopsis showed a high similarity to the protein databank structures of PDB ID: 1QZ6A, 2VYPA, 1QZ5A,

Table 5. Quality and Properties of Nests resulted by ProFunc Nest analysis

| Nest	Score	Residue range	Residue	Ramachandran region	Residue conservation
1	2.00	Ser147-Arg149	Ser147, Gly148, Arg149	right left –	1.0 1.0 1.0
2	2.00	Arg198-Tyr200	Arg198, Gly199, Tyr200	right left –	1.0 1.0 1.0
3	2.00	Asp246-Gln248	Asp246, Gly247, Gln248	right left –	1.0 1.0 1.0
4	2.00	Ile269-Met271	Ile269, Gly270, Met271	right left –	1.0 1.0 1.0
5	1.00	Leu96-Val98	Leu96, Arg97, Val98	right left –	1.0 1.0 1.0
6	1.00	Gly310-Ala312	Gly310, Ile311, Ala312	left right right	1.0 1.0 1.0
7	0.88	Phe129-Ala131	Phe129, Asn130, Ala131	right left –	1.0 1.0 0.63
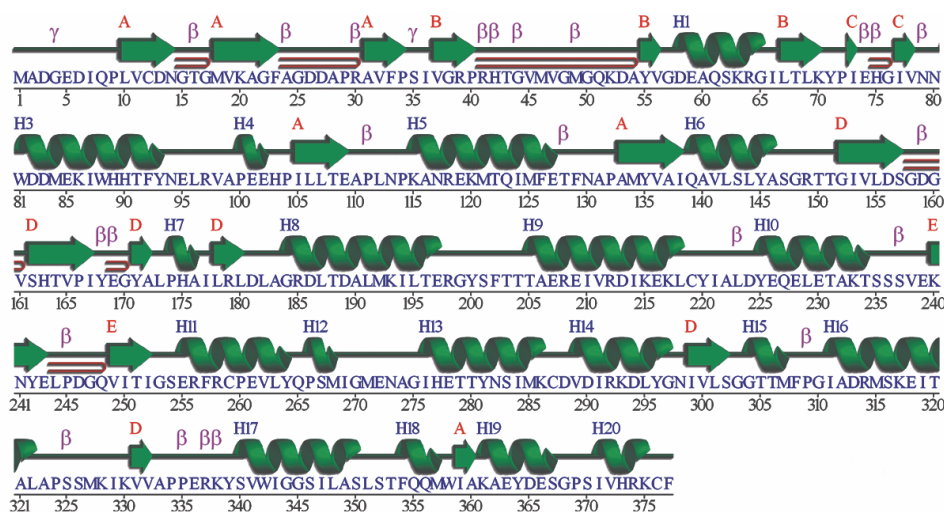
Fig. 8. Representation of the secondary structural topology of Actin-1 protein of *Arabidopsis thaliana*. Helices are labeled as (H1, H2...); Strands with their sheets are labeled as (A, B...); beta turns are labeled as β; and gamma turns are labeled as γ

1IJJA, 2W49D and 1MVW1. The degree of the matching fold similarity between the structures (Actin-1 of Arabidopsis and target) was signified by a high Z-score. The resulting Q-score, Z-score, number of SSE and RMSD between the C-alpha atoms of the matched residues and the sequence identity indicate a significance of the match and are summarized in Table 6. The putative binding sites such as jaspisamide A, myxobacterial rhizopodin, kabiramide C and latrunculin A are indicated in the modeled Actin-1 protein of Arabidopsis.

Furthermore, the ligand binding potency of the modeled Actin-1 was predicted using Galaxy web server (Soding, 2005; Joo et al., 2008; Pei, Kim and Grishin, 2008; Park et al., 2011; Ko et al., 2012a; Park and Seok,

Table 6. Predicted match fold in modeled Actin-1 structure of Arabidopsis

Hit	Q-score	Z-score	No. SSE	RMSD [Å]	Seq. id	PDB entry	Name
1	0.894	23.4	28	0.71	88.3%	*1qz6A*	structure of rabbit actin in complex with jaspis-amide A
2	0.880	23.4	29	0.67	88.5%	*2vypA*	rabbit-muscle g-actin in complex with myxobacterial-rhizopodin
3	0.873	22.7	28	0.87	88.3%	*1qz5A*	structure of rabbit actin in complex with kabir-amidE C
4	0.859	20.4	24	0.81	88.4%	*1ijjA*	the X-ray crystal structure of the complex between rabbit skeletal muscle actin and latrunculin A at 2.85 a resolution
5	0.839	22.8	28	0.99	86.3%	*2w49D*	isometrically contracting insect asynchronous flight muscle
6	0.839	22.8	28	0.99	86.3%	*1mvw1*	molecular models of averaged rigor crossbridges from tomograms of insect flight muscle
7	0.245	9.3	14	2.65	20.6%	*1mwmA*	parm from plasmid r1 adp form
8	0.224	7.2	14	2.86	20.0%	*1mwkA*	parm from plasmid r1 apo form
9	0.222	9.7	17	2.35	14.9%	*1e4fT*	ftsa (apo form) from thermotoga maritime
10	0.221	7.6	13	3.07	20.6%	*3js6A*	crystal structure of apo psk41 parm protein

2012; Ko, Park and Seok, 2012b). The results reported in the above-listed literature suggested three potential binding sites in the modeled protein structure such as ATP binding site [residues: 15G, 16T, 17G, 18M, 20K, 158G, 159D, 160G, 161V, 184G, 212R, 215K, 216E, 303G, 304G, 305T, 307M, 308F and 338K], 4-methyl-histidine binding site [residues: 16T, 73I, 74E, 75H, 76G, 77I, 160G and 161V] and BISTRAMIDE A binding site [residues: 135Y, 137A, 138I, 141V, 142L, 145Y, 148G, 149R, 170G, 171Y, 172A, 174P, 347I, 348L, 353T, 357M and 377F]. This prediction also reflects the bottom of the cleft generated by the two domains (I and II) of the

modeled Actin-1 as the possible ATP-binding site and governs the nucleotide-dependent conformational change similar to the template structure (1NLV-A).

Phylogenetic studies

The representative UniProt protein sequences, showing a similarity with the Actin-1 protein sequence of Arabidopsis, were retrieved on the basis of the ProFunc result and the evolutionary relationship between these sequences was analyzed by applying 1000 iteration boot-strapping steps as shown in Figure 9.
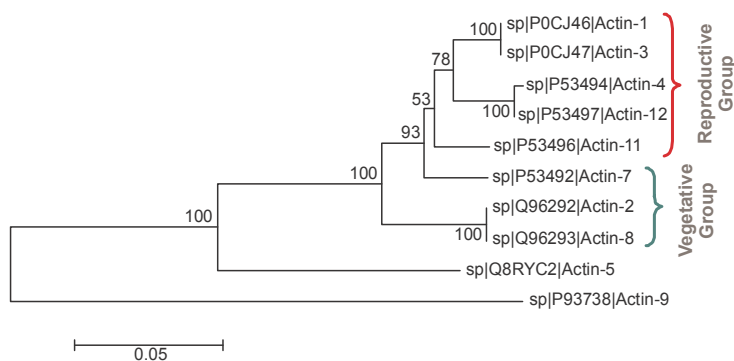

Fig. 9. Neighbor-Joining tree deduced from Actin-1 protein sequences of *Arabidopsis thaliana.* Bootstrap confidence values obtained with 1000 re-samplings are given at the branch point. Entries from this work are represented as: UNIPROT ID and protein name. Bar, 0.05 shows the substitutions per nucleotide position

The vegetative and the reproductive class of Actin proteins were distinctly divided considering the boot strap values as represented in Figure 9. There, Actin-1, -3, -4, -11 and -12 represent a reproductive group while Actin-2, -7 and -8 represent the vegetative group. The classification in Figure 9 indicates the number of non-conservative amino acid substitutions, among the actin proteins of Arabidopsis that may affect the protein-protein interactions.

Conclusions

Actin-1 is a highly conserved protein involved in various types of cellular motility and is ubiquitously expressed in all eukaryotic cells. It is an essential component of cell cytoskeleton, and plays an important role in cytoplasmic streaming, cell shape determination, cell division, organelle movement and extension growth. Misexpression of the pollen-specific reproductive Actin-1 isovariant in vegetative tissues alters actin polymerization and F-actin organization and thereby dramatically affects plant development and morphogenesis in *A. thaliana*. Understanding the functional biology of Actin-1 protein of *A. thaliana* may aid in locating the putative functional sites and understanding the phylogeny of Actin-1 protein. In this study, the structure of Actin-1 of Arabidopsis was solved using the comparative protein modeling approach. This report suggests the location of binding sites for jaspisamide A, myxobacterial rhizopodin, kabiramide C, latrunculin A as well as the enzymatically active domain and ATP binding domains in the modeled Actin-1 structure. The present study has explored significant motifs and well-conserved functional active regions in the modeled structure of Actin-1. Actin-1 shows a high degree of structural conservation from non-vertebrates to higher eukaryotes. In addition, preliminary phylogenetic studies indicate that the sequence of Actin-1 of Arabidopsis belonging to the family Brassicaceae shows higher sequence similarity to the actin proteins from legumes than that from other flowering plant families (data not shown). The predicted binding sites in Actin-1 will constitute a fundamental aspect for further protein-ligand interaction studies. The data generated from the present study is important for further studies on plant Actin in general, and that of *Arabidopsis thaliana* in particular. Again, the final simulated atomic coordinates of Actin-1 protein will be appropriate for protein docking studies as well as for modern *in silico* protein characterization. Further studies are needed to gain more insight in to the probable function of modeled Actin-1 protein.

References

Altshul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. (1990) *Basic local alignment search tool.* J. Mol. Biol. 215: 403-410.

Bailey T.L., Elkan C. (1994) *Fitting a mixture model by expectation maximization to discover motifs in biopolymers.* In: *Proceedings of the 2nd International Conference on Intelligent Systems for Molecular Biology*, Menlo Park, California 28: 36.

Bermen H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. (2000) *The protein data bank.* Nucl. Acids Res. 28: 235-242.

Chen C.Y., Wong E.I., Vidali L., Estavillo A., Hepler P.K., Wu H.M., Cheung A.Y. (2002) *The Regulation of Actin Organization by Actin-Depolymerizing Factor in Elongating Pollen Tubes.* Plant Cell 14: 2175-2190.

Cheung A.Y., Chen C.Y., Glaven R.H., de Graaf B.H., Vidali L., Hepler P.K., Wu H.M. (2002) *Rab2 GTPase Regulates Vesicle Trafficking between the Endoplasmic Reticulum and the Golgi Bodies and Is Important to Pollen Tube Growth.* Plant Cell 14: 945-962.

Clarke S.R., Staiger C.J., Gibbon B.C., Franklin-Tong V.E. (1998) *A potential signaling role for profilin in pollen of Papaverrhoeas.* Plant Cell 10: 967-979.

Colovos C., Yeates T.O. (1993) *Verification of protein structures: patterns of non-bonded atomic interactions.* Protein Sci. 2: 1511-1519.

Cristobal S., Zemla A., Fischer D., Rychlewski L., Elofsson A. (2001) *A study of quality measures for protein threading models.* BMC Bioinformatics: 2:5.

Dong C.H., Xia G.X., Hong Y., Ramachandran S., Kost B., Chua N.H. (2001) *ADF Proteins Are Involved in the Control of Flowering and Regulate F-Actin Organization, Cell Expansion, and Organ Growth in Arabidopsis.* Plant Cell 13: 1333-1346.

Eisenberg D., Luthy R., Bowie J.U. (1997) *VERIFY3D: Assessment of protein models with three dimensional profiles.* Meth. Enzymol. 277: 396-404.

Franklin-Tong V.E. (1999) *Signaling and the modulation of pollen tube growth.* Plant Cell 11: 727-738.

Fu Y., Li H., Yang Z. (2002) *The ROP2 GTPase Controls the Formation of Cortical Fine F-Actin and the Early Phase of Directional Cell Expansion during Arabidopsis Organogenesis.* Plant Cell 14: 777-794.

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

Hess B., Bekker H., Berendsen H.J.C., Fraaije J.G.E.M. (1997) *LINCS: A linear constraint solver for molecular simulations.* J. Comput. Chem. 18: 1463-1472.

Hess B., Kutzner C., Spoel D., Lindahl E. (2008) *GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation.* J. Chem. Theory. Comput. 4: 435-447.

Hooft R.W.W., Vriend G., Sander C., Abola E.E. (1996) *Errors in protein structures.* Nature 381: 272-272.

Joo K., Lee J., Seo J.H., Lee K., Kim B.G., Lee J. (2008) *All-atom chain-building by optimizing MODELLER energy function using conformational space annealing.* Proteins 75: 1010-1023.

Kandasamy M.K., McKinney E.C., Meagher R.B. (2002) *Functional Non equivalency of Actin Isovariants in Arabidopsis.* Mol. Biol. Cell. 13: 251-261.

Ko J., Park H., Heo L., Seok C. (2012a) *GalaxyWEB server for protein structure prediction and refinement.* Nucl. Acids Res. 40: W294-W297.

Ko J., Park H., Seok C. (2012b) *GalaxyTBM: template-based modeling by building of reliable core structures and refinement of unreliable local regions.* BMC Bioinformatics 13: 198.

Krissinel E., Henrick K. (2004) *Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions.* Acta. Cryst. D60: 2256-2268.

Laskowski R.A., MacArthur M.W., Moss D.S., Thornton J.M. (1993) *PROCHECK: A program to check the sterochemical quality of protein structures.* J. Appl. Cryst. 26: 283-291.

Laskowski R.A., Watson J.D., Thornton J.M. (2005a) *Protein function prediction using local 3D templates.* J. Mol. Biol. 351: 614-626.

Laskowski R.A., Watson J.D., Thornton J.M. (2005b) *ProFunc: a server for predicting protein function from 3D structure.* Nucl. Acids Res. 33: W89-W93.

Li H., Lin Y., Heath R.M., Zhu M.X., Yang Z. (1999) *Control of pollen tube tip growth by a RopGTPase-dependent pathway that leads to tip-localized calcium influx.* Plant Cell 11: 1731-1742.

Li X.B., Fan X.P., Wang X.L., Cai L., Yang W.C. (2005) *The Cotton ACTIN1 Gene Is Functionally Expressed in Fibers and Participates in Fiber Elongation.* Plant Cell 17: 859-875.

Linding R., Jensen L.J., Diella F., Bork P., Gibson T.J., Russell R.B. (2003) *Protein disorder prediction; implications for structural proteomics.* Structure 11: 1453-1459.

Maiti R., Van Domselaar G.H., Zhang H., Wishart D.S. (2004) *SuperPose: a simple server for sophisticated structural superposition.* Nucl. Acids Res. 32: W590-W594.

McDowell J.M., Huang S., McKinney E.C., An Y.Q., Meagher R.B. (1996) *Structure and Evolution of the Actin Gene Family in Arabidopsis thaliana.* Genetics 142: 587-602.

McKinney E.C., Ali N., Traut A., Feldmann K.A., Belostotsky D.A., McDowell J.M., Meagher R.B. (1995) *Sequence based identification of T-DNA insertion mutations in Arabidopsis: actin mutants act2-1 and act4-1.* Plant J. 8: 613-622.

Pal D., Suhnel J., Weiss M.S. (2002) *New principles of protein structure: nests, eggs – and what next?* Angew. Chem. Int. Ed. 41: 4663-4665.

Park H., Ko J., Joo K., Lee J., Seok C., Lee J. (2011) *Refinement of protein termini in template-based modeling using conformational space annealing.* Proteins: Struct. Funct. Bioinfor. 79: 2725-2734.

Park H., Seok C. (2012) *Refinement of unreliable local regions in template-based protein models.* Proteins: Struct. Funct. Bioinfor. 80: 1974-1986.

Pei J., Kim B.H., Grishin N. (2008) *PROMALS3D: a tool for multiple protein sequence and structure alignments.* Nucl. Acids Res. 36: 2295-2300.

Porter C.T., Bartlett G.J., Thornton J.M. (2004) *The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data.* Nucl. Acids Res. 32: D129-D133.

Quevillon E., Silventoinen V., Pillai S., Harte N., Mulder N., Apweiler R., Lopez R. (2005) *InterProScan: protein domains identifier.* Nucl. Acids Res. 1: W116-W120.

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

Sali A., Matsumoto R., McNeil H.P., Karplus M., Stevens R.L. (1993b) *Three-dimensional models of four mouse mast cell chymases, identification of proteoglycan-binding regions and protease-specific antigenic epitops.* J. Biol. Chem. 268: 9023-9034.

Sali A., Overington J.P. (1994) *Derivation of rules for comparative protein modeling from a database of protein structure alignments.* Protein Sci. 31: 1582-1596.

Sali A., Pottertone L., Yuan F., Vlijmen V.H., Karplus M. (1995) *Evaluation of comparative protein modeling by MODELLER.* Proteins 23: 318-326.

Scott W.R.P., Hunenberger P.H., Tironi I.G., Mark A.E., Billeter S.R., Fennen J., Torda A.E., Huber T., Kruger P., Gunsteren W.F. (1999) *The GROMOS Biomolecular Simulation Program Package.* J. Phys. Chem. A. 103: 3596-3607.

Sippl M.J. (1993) *Recognition of Errors in Three-Dimensional Structures of Proteins.* Proteins 17: 355-362.

Soding J. (2005) *Protein homology detection by HMM-HMM comparison.* Bioinformatics 21: 951-960.

Tamura K., Dudley J., Nei M., Kumar S. (2007) *MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.* Mol. Biol. Evol. 24: 1596-1599.

Thompson J.D., Higgins D.G., Gibson T.J. (1994) *CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.* Nucl. Acids Res. 22: 4673-4680.

Vorobiev S., Strokopytov B., Drubin D.G., Frieden C., Ono S., Condeelis J., Rubenstein P.A., Almo S.C. (2003) *The structure of non vertebrate actin: implications for the*

ATP hydrolytic mechanism. Proc. Natl. Acad. Sci. USA 100: 5760-5765.

Watson J.D., Milner-White E.J. (2002a) *The conformations of polypeptide chains where the main-chain parts of successive residues are enantiomeric. Their occurrence in cation and anion-binding regions of proteins.* J. Mol. Biol. 315: 183-191.

Watson J.D., Milner-White E.J. (2002b) *A novel main-chain anion-binding site in proteins: the nest. A particular combination of phi, psi values in successive residues gives rise to anion-binding sites that occur commonly and are found often at functionally important regions.* J. Mol. Biol. 315: 171-82.

Wiederstein M., Sippl M.J. (2007) *ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins.* Nucl. Acids Res. 35: W407-W410.

Yang Z. (1998) *Signaling tip growth in plants.* Curr. Opin. Plant Biol. 1:525-30.