International Journal of Pharmacy and Biological Sciences-IJPBS™ (2019) 9 (2): 269-277
Online ISSN: 2230-7605, Print ISSN: 2321-3272
Research Article | Biological Sciences | Open Access | MCI Approved
UGC Approved Journal

In-Silico Analysis of Structural and Functional Aspects of Salt Responsive Protein of SKP-1 Like Protein 1A

G. Anbarasi[1]*, M. P. Arulmoorthy[2] and R. Suresh[2]
[1*]PG and Research Department of Biotechnology, Kongunadu Arts and Science College, Affiliated to Bharathiyar University, Coimbatore - 641 029, Tamilnadu, India.
[2]CAS in Marine Biology, Annamalai University, Parangipettai - 608 502, Tamilnadu, India.

Received: 15 Jan 2019 / Accepted: 12 Mar 2019 / Published online: 1 Apr 2019
Corresponding Author Email: anbarasigbio@gmail.com

Abstract
Abiotic stresses limiting plant growth and productivity. By modulating the amount and activity of regulatory proteins, ubiquitination plays a central role in regulating the transcriptional changes required for adaption to abiotic stresses. S-phase kinase-associated protein 1 (SKP1) is a component of a Skp1-Cullin1-F-box (SCF) complex that facilitates ubiquitin-mediated protein degradation in stress-related signaling and response mechanism. The present study has been carried out to predict the structure and function of a salt responsive protein of SKP1-like protein1A. The uncharacterized protein analyzed in the present study showed conserved domain characteristics of sequence resembles S-phase kinase-associated protein 1 (SKP1) family. The protein exhibited a maximum number of random coils (30.72%)) with alpha helix (63.86%) and extended strands (5.42%)) as secondary structural elements. Three-Dimensional modeling was carried out to elucidate its structure and its active sites. These results aid to the experimental data and help to built up a complete view of SKP1-like protein1A role in plant stress response.

Keywords
SKP1, Ubiquitin protein degradation, stress protein, Salt tolerance, proteomics.

*****

INTRODUCTION
In silico analysis of different gels provide a complete catalogue that exists between different sample protein patterns [1]. Protein identification and analysis is usually based on *in silico* techniques that match the peptide sequences that paves the way in metabolomics, transcriptomic and other large pomics techniques into systems biology [2]. There are tremendous opportunities for molecular biologists to define the protein nature that the transducer genetic information and energy transfer that obtains from plants. These proteomics methods are very much important for the study of different components in the plant that states plant function [3].

When carrying out analysis of protein sequences, the aim is to find out as much information as possible about potential relationships with other sequences as well as characterizing their physiochemical properties. The first step usually involves comparing

the protein sequence against a non-redundant protein sequence database by using Blast [4] or Fasta [5], which will reveal which sequence(s) are similar to the query sequence alone. To obtain further information about a proteins specific function, searches against secondary databases also known as pattern or signature databases are necessary. When such searches return significant matches or hits, these results help in the assignment of a particular function or functional domain to the query protein [6].

Proteins are important biological polymers formed from building blocks called amino acids. The three-dimensional structure and biological activity of proteins depend on the physicochemical properties of their constituent amino acids. The primary structure identifies a protein unambiguously, determines its chemical and biological characteristics, and specifies the higher levels of protein structure [7].

Prediction of secondary structures is a fundamental basis for protein structure prediction. Protein structure determination and prediction has been a focal research subject in the field of bioinformatics due to the importance of protein structure in understanding the biological and chemical activities of organisms [8]. The term "homology modeling", also called comparative modeling or template-based modeling (TBM), refers to modeling a protein 3D structure using a known experimentally determined structure of a homologous protein as a template. Protein modeling is the only way to obtain structural information if experimental techniques fail. Many proteins are simply too large for NMR analysis and cannot be crystallized for X-ray diffraction [9].

In the present study, *In silico* characterization of amino acid sequence of SKP1-like protein 1A protein was carried out. Secondary structures of the protein based on their residues were predicted and classified. Type of residues and their physical properties have been estimated. By profiling the compositional analysis of the proteins the charge distribution, atomic composition, extinction coefficients, instability index and aliphatic index were predicted. The present investigation implies homology modeling to deduce a three dimensional structure of the uncharacterized protein under study, followed by validation and comparative modeling of the obtained structure for its conformational stability and further biological analysis.

MATERIALS AND METHODS

Sequence analysis

The sequences that are retrieved from MALDI analysis were selected for *in silico* analysis. The sequences were compared for detecting homologous sequences found in databases using Basic Local Alignment Search Tool (BLAST) (4). BLAST from NCBI was used to compare the query sequence with the database sequence to find its homologues.

MOTIF search

The motifs were identified using the tool Motif Search [10].

Conserved domain search

Conserved domain search was performed using CDD [11]. Enter a protein or nucleotide query as an accession or GI number (e.g., AAC50285 or 463989) or as a sequence in FASTA format to identify the protein's conserved domains and therefore its putative function.

InterPro based protein signature recognition analysis

The InterPro project home page is at http://www.ebi.ac.uk/interpro. InterProScan can take either nucleotide or protein sequences in a recognized sequence format (such as raw, FASTA or EMBL). It will reformat and, if necessary, translates the sequences before beginning its search tasks. If raw format (free text) is used, it will be given the name "Sequence_n" by default, where n is the order in which it appeared in the input [12].

ProtParam based physical and chemical computation in protein sequences

Using the primary sequence, the physicochemical properties of the protein were calculated with the aid of the tool ProtParam. The computed parameters include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity [13].

Secondary structure prediction

The GOR IV server analyzes sequences to predict alpha helix, beta sheet, turn or random coil secondary structure at each position based on 17-amino-acid sequence windows [14].

Homology modeling

For homology modeling, initially a suitable template was searched using Protein data bank [4], SWISS model template library [15] and PSI-BLAST [16]. Amino acid sequence alignment of target and template proteins and rough 3D models (03 models) were constructed by using homology modeling servers such as Swiss models [14].

Model structure validation

All the generated structures of the protein model were subjected to a series of tests for testing its internal consistency and reliability. Backbone conformations were evaluated by the inspection of the Psi/Phi Ramachandran plot obtained from PROCHECK analysis [17]. The analyses can be generated by clicking on the PROCHECK button at the top of the PROCHECK summary page; and can get to this page from the PDB sum page of the structure in question either by clicking on PROCHECK in the index on the left hand side of the page, or else by clicking on the Ramachandran plot on the right.

Structure visualization

The modeled tertiary structure was visualized using Rasmol. It reads molecular structure files from the Protein Data Bank (PDB) in .pdb extension. The web server is available at http://www.openrasmol.org.

RESULTS AND DISCUSSION

Sequence analysis SKP1-like protein 1A

In silico analysis of the obtained sequence was carried out. BLAST homology searches showed that this sequence shares 100% identity with PREDICTED: SKP1-like protein 1A (*Sesamum indicum*).

MOTIF search SKP1-like protein 1A revealed that about 3 motifs were present in the test sequence such as SKP1, SKP1_POZ and Ribosomal _L7Ae. Positions and E-values are given in (Fig 1).

Conserved domain search revealed that, sequence resembles S-phase kinase-associated protein 1 (SKP1) family. The domain identified in the protein sequence was positioned at 89 to 166 residues (Fig 2).

In the present study, the protein signature analysis for the SKP1-like protein1A sequences of the study species with highlighted domains corresponding to the matches in the InterPro database was obtained (Fig 3). This result again confirms that uncharacterized protein sequence belongs to SKP1 super family and containing three domains of SKP1 component POZ domain, POZ domain and SKP1 component dimerisation.

Result analysis of the protein PREDICTED: SKP1-like protein 1A showed the motif and conserved domain characteristics of S-phase kinase-associated protein 1 (SKP1) super family. Generally, plant SKP1-like family proteins, components of the Skp1-Cullin1-F-box SCF complex E3 ligases, are involved in the regulation of plant development and stress responses [18]. The relevance of the ubiquitin-dependent protein degradation in stress-related signalling and response mechanism [19,20,21].

Physicochemical properties of the protein

Prot Param results exhibit the physicochemical parameters of uncharacterized protein [12]. There are 166 amino acids in the sequence, its molecular weight was 18347.6 and theoretical pI was 4.53. The maximum number of amino acids present in the sequence were Alanine (12.7%) followed by Glutamine (10.8%) and Aspargene (9%) (Table 1). The total numbers of negatively charged residues (Aspartic acid + Glutamic acid) were thirty-three while the total numbers of positively charged residues (Arginine + Lysine) were nineteen. Atomic composition computed as C (802), H (1281), N (209), O (267) and S (7). Thus $C_{802}H_{1281}N_{209}O_{267}S_7$ has been arrived as the molecular formula for the SKP1like protein A protein. The aliphatic index was calculated as 83.55. The instability index is computed to be 46.64 which classified the protein as unstable. The grand average of hydropathicity (GRAVY) was calculated to be -0.336. Amino acid composition of predicted SKP1 like Protein 1A was showed in Table 2.

Secondary structure prediction

The secondary structural analysis of the protein was done (Fig 4) and alpha helix (Hh) was found to be most frequent (63.86%), followed by random coil (30.72%). Extended strand (Ee) was found to be least frequent (5.42%) (Table 2). The dominance of the coiled regions indicates the high level of conservation and stability of the protein structure [22].

Information about amino acid composition would also be useful for refining the analysis of truncated hydrophobic clusters that only cover a limited part of the associated regular secondary structure. Amino acids in the vicinity of these hydrophobic clusters, such as A, C and T that can substitute for strong hydrophobic residues, might indicate the overflowing of the hydrophobic cluster by the associated regular secondary structure.

The modeled stress induced protein of *Capnocytophaga canimorsus Cc5* exhibited a maximum number of random coils (47.06%) with alpha helix (36.03%) and extended strands (16.91%) as secondary structural elements (23). In this study alpha helix (Hh) was found to be most frequent (63.86%), followed by random coil (30.72%) and extended strand (Ee) was found to be least frequent (5.42%). However, this study protein contains more structural elements.

Homology modeling

Template search with BLAST was performed against PDB and SWISS-model template library. A total of 32 templates were found. Suitable tree templates of

PDB ID: 3ogm (*Arabidopsis thaliana*), 2ass (*Homo sapiens*), 3ogl (*Arabidopsis thaliana*), were selected on the basis of sequence similarity with a resolution of 78% with 3.34A^{0,} 60.8% with 3A^{0} and 79% with 3.18 A^{0} respectively. For each identified templates, the template quality was predicted from the features of the target-template alignment and rough 3D models were built for three templates using SWISS modeler.

Structure validation

The models were validated in PROCHECK. The formation of helix-helix contacts is crucial for membrane protein folding. Residues involved in helix-helix contacts are therefore evolutionary conserved. Ramachandran plot shows the phi-psi torsion angles for all residues in the structures of the three models for target protein of SKP1-like protein A. The coloring/shading on the plot represents the different regions [24]. The darkest areas shown in red correspond to the "core" regions representing the most favorable combinations of phi-psi values (Fig 5).

The percentage of residues in the “core” regions is one of the better guides to stereochemical quality (Table 3). The plot shows separate Ramachandran plots for each of the 20 different amino acid types. Shading on each plot indicates how favorable each region on the plot is; the darker the shade the more favorable the region. The numbers in brackets, following each residue name, show the total number of data points on that graph (Fig 6 a-c). The red numbers above the data points are the reside-numbers of the residues lying in unfavorable regions of the plot. This gives a visualization of which regions appear to have consistently poor or unusual geometry and which have more normal geometry [15].

The PROCHECK analysis of uncharacterized protein generated from heat responsive showed 65.2% residues in most favored region (A, B, L), 30.3% residues in additional allowed region and 4.5% residues in disallowed region [25]. Ramachandran plot for salt stress responsive transcription factor SsMYB2R protein (*Saccharum sportuneum*) obtained through PROCHECK founded the residues of protein in the core region was 78.1% [26]. In the present study, percentage of residues in the core region of protein sample was 79.9%, 80.1% and 82.8% for the three models. Ramachandran plot results highlighted that the SKP1-like protein1A protein sequences of model 3 showed high stereochemical quality when compared to that of model 1and model 2 sequences. When comparing the results with previous reports of stress responsive proteins such as uncharacterized protein and SsMYB2R protein the percentage of stereochemical quality of this SKP1-like protein 1A was found to be comparatively high.

Tertiary structure of SKP1-like 1A

Based on the result of sequence similarity, resolution value and the validation report, model 3 (PDB ID: 3ogl (*Arabidopsis thaliana*)) was selected as a best template for the target sequence of SKP1-like protein1A and finally tertiary structure of the modeled protein visualized using Rasmol (Fig 7).

Table 1. Amino acid composition of SKP1 like protein 1A

Amino acid	No. of residues	% of residues
Ala (A)	21	12.7%
Arg (R)	4	2.4%
Asn (N)	7	4.2%
Asp (D)	15	9.0%
Cys (C)	3	1.8%
Gln (Q)	4	2.4%
Glu (E)	18	10.8%
Gly (G)	4	2.4%
His (H)	2	1.2%
Ile (I)	11	6.6%
Leu (L)	11	6.6%
Lys (K)	15	9.0%
Met (M)	4	2.4%
Phe (F)	7	4.2%
Pro (P)	5	3.0%
Ser (S)	10	6.0%
Thr (T)	11	6.6%
Trp (W)	1	0.6%
Tyr (Y)	2	1.2%

Val (V)	11	6.6%
Pyl (o)	0	0.0%
Sec (U)	0	0.0%

Table 2. Details for Secondary structure of SKP1 like protein 1A (using GOR IV)

Structural elements Number of residues	Number of residues	Percentage of residues
Alpha helix (Hh)	106	63.86 %
310 helix (Gg)	0	0.00 %
Pi helix (Ii)	0	0.00 %
Beta bridge (Bb)	0	0.00 %
Extended strand (Ee)	9	5.42 %
Beta turn (Tt)	0	0.00 %
Bend region (Ss)	0	0.00 %
Random coil (Cc)	51	30.72 %
Ambigous states (?)	0	0.00 %
Other states	0	0.00 %

Table 3. Percentage of residues falling in the core region of Ramachandran's plot for the three template models of 1, 2 and 3 (PDB ID 3ogm, 2ass and 3ogl).

Percentage parameters	Model 1(3ogm)	Model 2 (2ass)	Model 3 (3ogl)
% of residue in most favored regions	79.9	80.1	82.8
% of residue in the additionally allowed zones	19.2	17.4	16.5
% of residue in the generously regions	0.2	2.5	0.7
% of residue in disallowed regions	0.0	0.0	0.0
% of non-glycine and non-proline residues	100	100	100

Figure 1. Predicted motif search of SKP1- like protein 1A

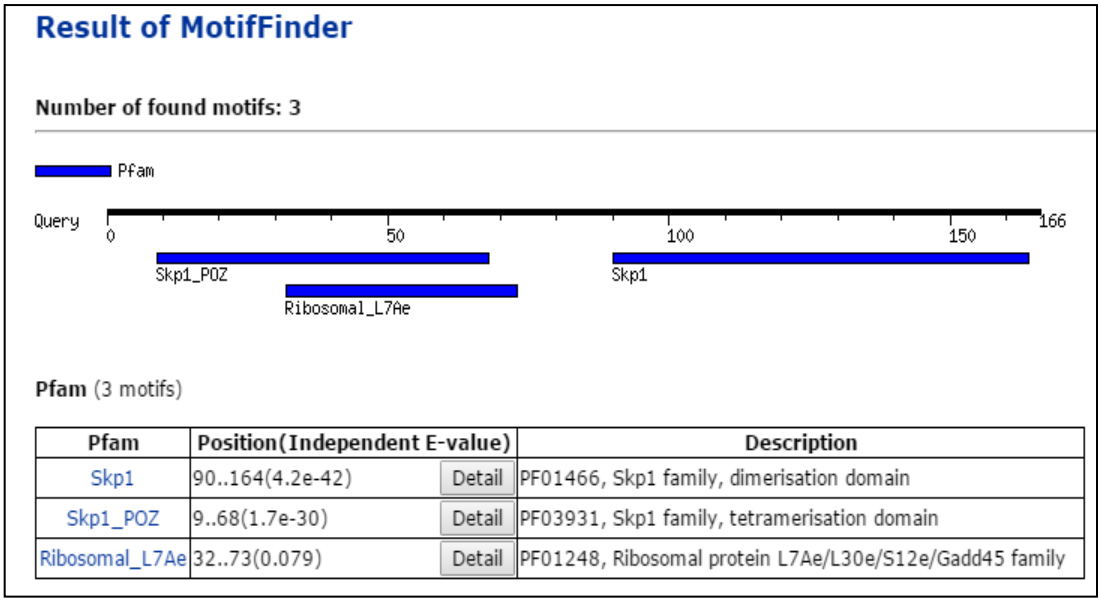
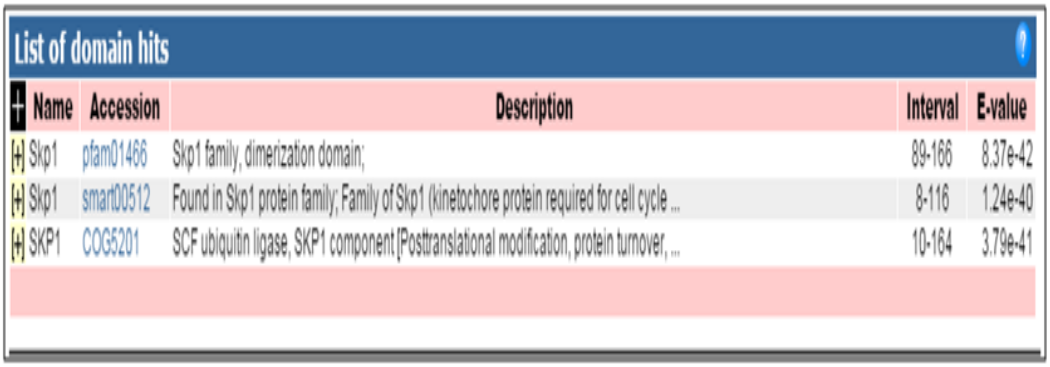


Figure 2. Conserved domain search (CDD) using BLAST for predicted SKP1-like protein 1A



List of domain hits

Name	Accession	Description	Interval	E-value
Skp1	pfam01466	Skp1 family, dimerization domain;	89-166	8.37e-42
Skp1	smart00512	Found in Skp1 protein family; Family of Skp1 (kinetochore protein required for cell cycle ...	8-116	1.24e-40
SKP1	COG5201	SCF ubiquitin ligase, SKP1 component [Posttranslational modification, protein turnover, ...	10-164	3.79e-41

Figure 3. Integrated protein signature analysis for predicted SKP1-like protein 1A

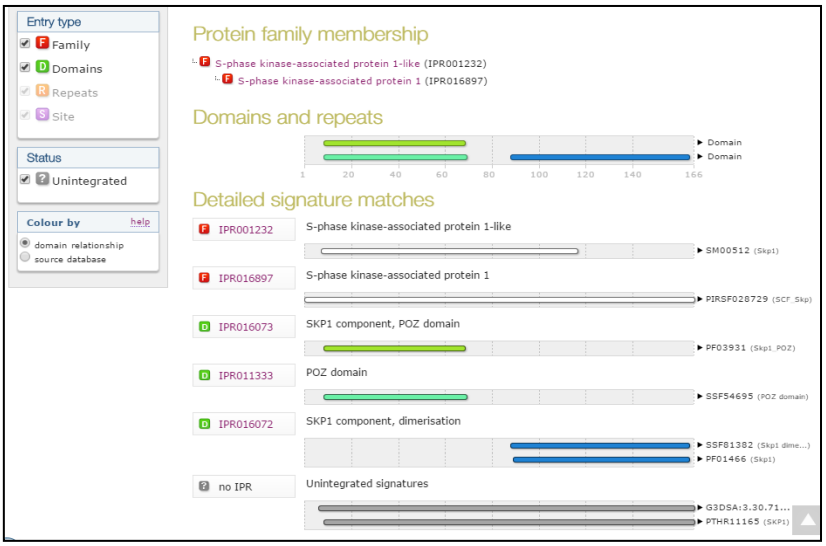
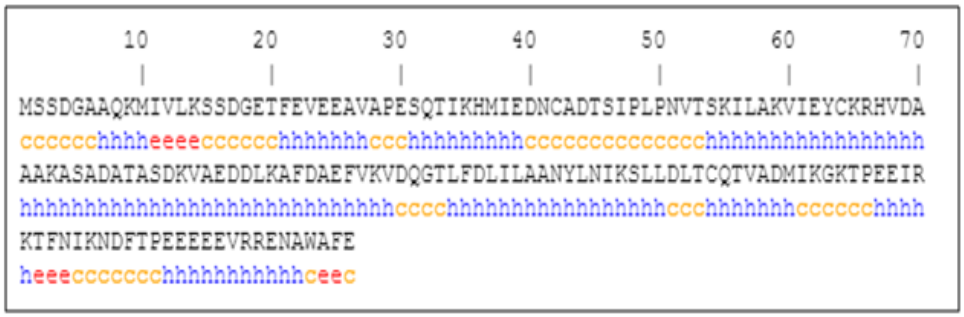


Figure 4. Secondary elements of the protein exhibited by GOR



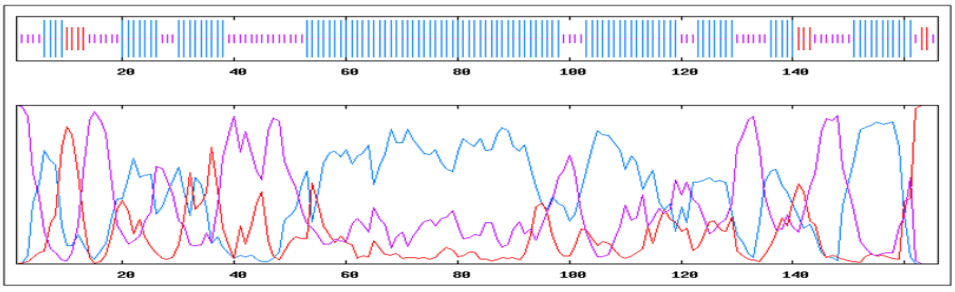


Figure 5. Ramachandran analysis of the backbone dihedral angles Psi (j) and Phi (s) for the three template models PDB ID 3ogm, 2ass and 3ogl validated with ProCheck program. Red region represents the most favored region, yellow = allowed region, light yellow = generously allowed region, white = disallowed region.

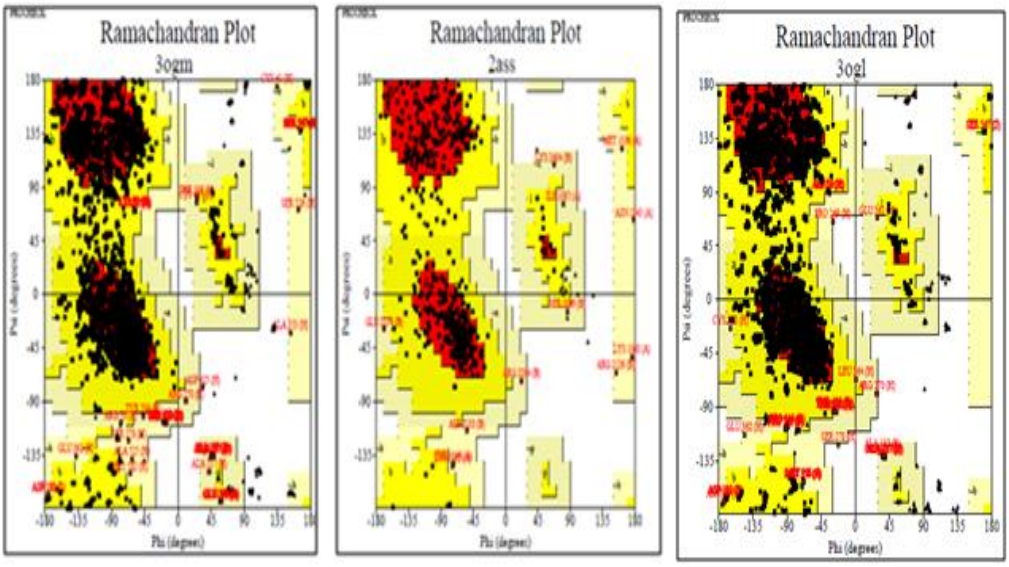


Figure 6a. The plot shows separate Ramachandran plots for different amino acid types for model 1 (3ogm).

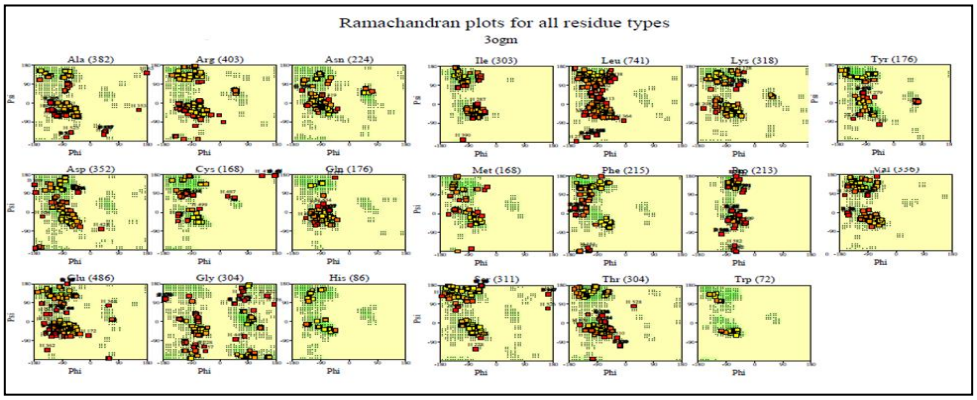


Figure 6b. The plot shows separate Ramachandran plots for different amino acid types for model 2 (2ass).

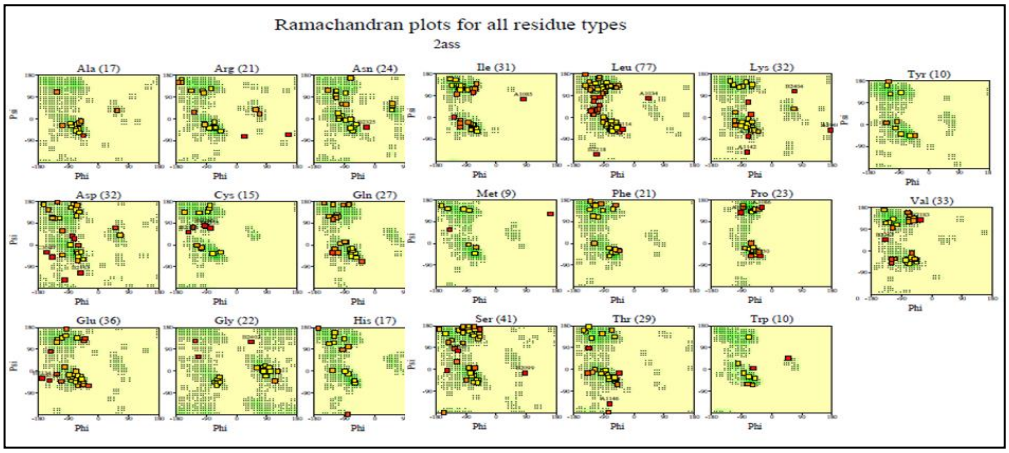


Figure 6c. The plot shows separate Ramachandran plots for different amino acid types for model 3 (3ogl)

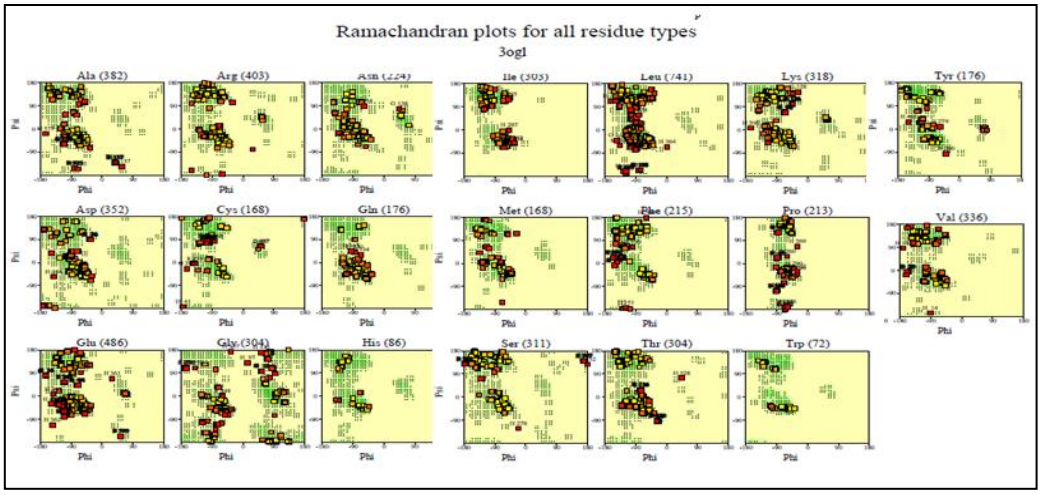
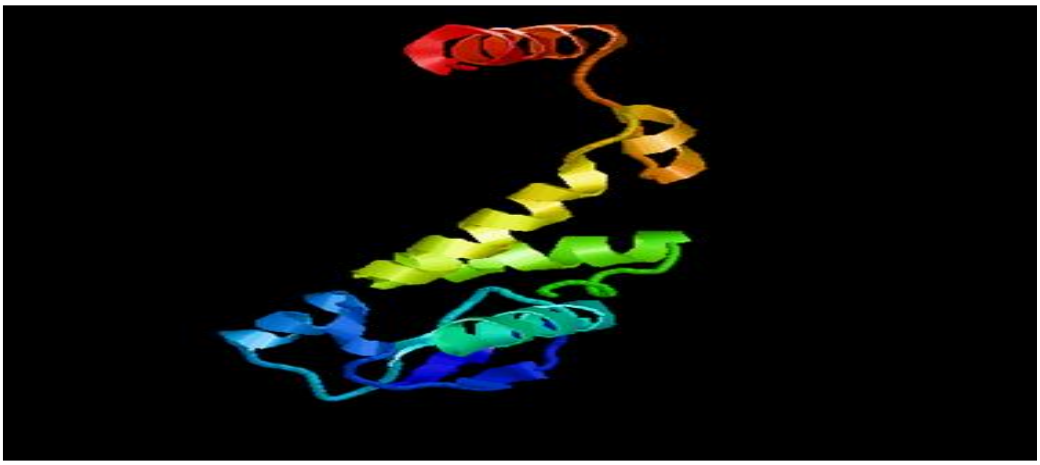


Figure 7. Tertiary structure view of modeled SKP1-like protein 1A



CONCLUSION

The sequence and structural analysis of S-phase kinase-associated protein1–like protein1A (SKP1-like protein1A) was carried out using bioinformatics tools. Sequence analysis of SKP1-like protein1A revealed the presence of highly conserved region of SKP1 protein family. The sequence analysis revealed that SKP1-like protein1A protein might be involved in ubiquitine dependent protein degradation in stress related signaling and response mechanism. Thus results aid to the experimental data and help to built up a complete view of SKP1-like protein1A protein role in plant stress response.

ACKNOWLEDGMENTS

The author is thankful to the PG and Research Department of Biotechnology, Kongunadu Arts and Science College and CAS Marine Biology, Annamalai University, Parangipettai, Cuddalore for providing necessary facility and support to carry out the research work successfully.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] Pirondini A, Visioli G, Malcevschi A, Marmiroli, NA. (2006). 2-D liquid-phase chromatography for proteomic analysis in plant tissues. J chromatography, 833(1): 91-100.

[2] Chen S, Harmon A.C. (2006) Advances in plant proteomics. Proteomics, 6(20):5504-5516.

[3] Roberts JK. (2002). Proteomics and a future generation of plant molecular biologists. In Functional Genomics, Springer Netherlands, 143-154.

[4] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang ZZ, Miller W, Lipman DJ. (1997) Gapped Blast and Psi-Blast: A new generation of protein database search programs. Nucleic Acids Res, 25: 3389-3402.

[5] Pearson WR, Lipman DJ. (1988). Improved tools for biological sequence comparison. Proc Natl Acad Sci. U.S.A, 85(2): 444–2448.

[6] Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. (2005) Inter Pro Scan: protein domains identifier. *Nucleic. Acids. Res*, 33: 116-120.

[7] Cozzone, A.J. (2002) Proteins: Fundamental Chemical Properties. Institute of Biology and Chemistry of Proteins, CNRS, Lyon, France, 1-10.
[8] Singh R, Xu J, Berger B. (2008) Global alignment of multiple protein interaction networks. Pac Symp Biocomput., 13: 303-314.
[9] Krieger E, Sander B, Nabuurs, GertVriend. Homology Modeling. Structural Bioinformatics. Edited by Philip E. Bourne and HelgeWeissig. Wiley-Liss, Inc.2003.
[10] Thakallapally R, Kibbe, W, Lang D, Korber B. Motifscan (2010): A Web-based Tool to Find HLA Anchor Residues in Proteins or Peptides. Theoretical Biology and Biophysics.
[11] Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K. (2011) CDD: A Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res., 39: 225-229.
[12] Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Bernard T, Binns D, Bork P, Burge S. (2012). InterPro in 2011: New developments in the family and domain prediction database. Nucleic Acids Res, 40: 306-312.
[13] Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. (2005) Protein identification and analysis tools on the expasy server. In: Walker JM (ed) The proteomics protocols handbook, Humana Press, New Jersey.
[14] Garnier J, Gibrat J, Robson B. (1996) GOR secondary structure prediction method version IV. Methods in Enzymology, R.F. Doolittle Ed, 266: 540-553.
[15] Arnold K, Bordoli L, Kopp J, Schwede T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling. Bioinformatics. 22:195-201.
[16] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PEc (2000). The protein data bank. Nucleic Acids Res. 28:235-242.
[17] Laskowski R, MacArthur M, Moss D, Thornton J. Procheck (2009): a program to check the stereochemical quality of protein structures. J Appl Cryst. 26: 283-291.
[18] Hotton SK, Callis J. (2008) Regulation of Cullin RING Ligases. Annu Rev Plant Biol. 59: 467–489.
[19] Ellis C, Turner JG, Devoto A. (2002) Protein complexes mediate signaling in plant responses to hormones, light, sucrose and pathogens. Plant Mol Biol. 50: 971–980.
[20] Zhang YY, Xie Q. (2007) Ubiquitination in abscissic acid-related pathway, J Int Plant Biol. 49: 87–93.
[21] Zhu J, Dong CH, Zhu JK. (2007) Interplay between cold-responsive gene regulation, metabolism and RNA processing during plant cold acclimation. Curr Opin Plant Biol., 10: 290–295.
[22] Neelamathi E, Vasumathi E, Bagyalakshmi S. 2009. Insilico prediction of structure and functional aspects of a hypothetical protein of *Neurospora crassa*. J Cell Tissue Res., 93:1889-1894
[23] Pilley HH. (2002) In-silico prediction of structural and functional aspects of a hypothetical protein of *capnocytophaga canimorsus Cc5.* J Adv Bioinfo Appl Res., 2(3): 206-210.
[24] Morris AL, MacArthur MW, Hutchinson EG, Thornton JM. (1992) Stereochemical quality of protein structure coordinates. Proteins., 12: 345-364.
[25] Padaria JC, Deepesh Bhatt, Koushik Biswas, Gagandeep Singh, Rajkumar Raipuria. (2013). *In-silico* prediction of an uncharacterized protein generated from heat responsive SSH library in wheat (*Triticum aestivum* L.). POJ. 6(2): 150-156.
[26] Kulkarni PA, Devaramuth RM. (2014) *In silico* 3D structure prediction SsMYB2R: A noval MYB transcription factor from *Saccharum spontaneum.* I J.biotech, 13: 437-447.